OM nucleic - nucleic search, using sw model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: May 4, 2004, 19:26:18 ; Search time 4456 Seconds
                                  (without alignments)
                                  11147.019 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 1146
Sequence: 1 atggctccgggcgaaagga........tgattacgctactcaagtga 1146

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
                 1: gb_ba:*
                 2: gb_htg:*
                 3: gb_in:*
                 4: gb_om:*
                 5: gb_ov:*
                 6: gb_pat:*
                 7: gb_ph:*
                 8: gb_pl:*
                 9: gb_pr:*
                 10: gb_ro:*
                 11: gb_sts:*
                 12: gb_sy:*
                 13: gb_un:*
                 14: gb_vi:*
                 15: em_ba:*
                 16: em_fun:*
                 17: em_hum:*
                 18: em_in:*
                 19: em_mu:*
                 20: em_om:*
                 21: em_or:*
                 22: em_ov:*
                 23: em_pat:*
                 24: em_ph:*
                 25: em_pl:*
                 26: em_ro:*
                 27: em_sts:*
                 28: em_un:*
                 29: em_vi:*
                 30: em_htg_hum:*
                 31: em_htg_inv:*
                 32: em_htg_other:*
                 33: em_htg_mus:*
                 34: em_htg_pln:*
                 35: em_htg_rod:*
                 36: em_htg_mam:*
                 37: em_htg_vrt:*
                 38: em_sy:*
                 39: em_htgo_hum:*
                 40: em_htgo_mus:*
                 41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1146	100.0	1146	9	AF503617	AF503617 Homo sapi
2	1144.4	99.9	1322	6	AX747674	AX747674 Sequence
3	1144.4	99.9	1322	9	AK092694	AK092694 Homo sapi
4	1142.8	99.7	2781	9	HSM804735	AL833422 Homo sapi
5	1141.2	99.6	2832	9	HSM803459	AL832152 Homo sapi
6	791.2	69.0	1128	10	AF504061	AF504061 Mus muscu
7	784.8	68.5	1285	10	AF336113	AF336113 Rattus no
8 c	667.4	58.2	152130	9	AP000856	AP000856 Homo sapi
9 c	667.4	58.2	153147	9	AF303733	AF303733 Homo sapi
10 c	667.4	58.2	186751	9	AC023344	AC023344 Homo sapi
11 c	419.8	36.6	174806	10	AC138715	AC138715 Mus muscu
12	419.8	36.6	174846	10	AC129212	AC129212 Mus muscu
13	419.8	36.6	186137	10	AC099736	AC099736 Mus muscu
14 c	373	32.5	73341	2	AC136853	AC136853 Rattus no
15	373	32.5	220702	2	AC114181	AC114181 Rattus no
16	227.2	19.8	145746	5	AL953858	AL953858 Zebrafish
17 c	213.2	18.6	177312	2	BX324201	BX324201 Danio rer
18	213.2	18.6	240809	2	BX571763	BX571763 Danio rer
19	205	17.9	261	6	AX884478	AX884478 Sequence
20	205	17.9	261	6	BD024088	BD024088 Sequence
21	205	17.9	265	6	BD075996	BD075996 5' EST of
22 c	72	6.3	7218	6	I66494	I66494 Sequence 14
23	56.8	5.0	10732	2	E32986	E32986 Gene encodi
24	56.4	4.9	1943	3	AY060993	AY060993 Drosophil
25	55	4.8	66497	2	AC015388	AC015388 Drosophil
26	55	4.8	171451	3	AC010561	AC010561 Drosophil
27	55	4.8	304419	3	AE003476	AE003476 Drosophil
28	53.4	4.7	863	3	AY113283	AY113283 Drosophil
29	52.4	4.6	22747	3	CEF36F2	Z81532 Caenorhabdi
30	52.2	4.6	2000	6	AX655393	AX655393 Sequence
31 c	51.8	4.5	39482	3	CET04A8	Z35663 Caenorhabdi
32	49.8	4.3	24604	3	DMC152A3	AL009194 Drosophil
33 c	49.8	4.3	67513	2	AC017770	AC017770 Drosophil
34	49.8	4.3	177724	3	AC104146	AC104146 Drosophil
35	49.8	4.3	193714	3	AC104149	AC104149 Drosophil
36	49.8	4.3	292830	3	AE003423	AE003423 Drosophil
37	46.6	4.1	125020	3	AF429315	AF429315 Homo sapi
38	45.6	4.1	256744	2	AC132718	AC132718 Rattus no
39 c	45.6	4.0	125020	9	AF429315	AF429315 Homo sapi
40 c	45.4	4.0	195539	2	AC107665	AC107665 Mus muscu
41 c	44.8	3.9	173599	2	AC141388	AC141388 Rattus no
42 c	43.8	3.8	231676	2	AC134063	AC134063 Rattus no
43 c	43.4	3.8	121402	10	AL929064	AL929064 Mouse DNA
44 c	43.4	3.8	159560	10	AL929036	AL929036 Mouse DNA
45 c	43.2	3.8	195729	2	AC132940	AC132940 Homo sapi

ALIGNMENTS

RESULT 1
AF503617
LOCUS   AF503617   1146 bp   mRNA   linear   PRI 01-JUL-2002
DEFINITION   Homo sapiens STARS mRNA, complete cds.
ACCESSION   AF503617
VERSION   AF503617.1 GI:20530823
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1146)
AUTHORS   Arai,A., Spencer,J.A. and Olson,E.N.
TITLE   STARS, a Striated Muscle Activator of Rho Signaling and Serum
             Response Factor-dependent Transcription

Pred. No. is the number of results predicted by chance to have a

```
JOURNAL      J. Biol. Chem. 277 (27), 24453-24459 (2002)
MEDLINE      22086201
PUBMED       11983702
REFERENCE    2 (bases 1 to 1146)
AUTHORS      Arai,A., Spencer,J.A. and Olson,E.N.
TITLE        Direct Submission
JOURNAL      Submitted (17-APR-2002) Molecular Biology, University of Texas
             Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
             75390-9148, USA
FEATURES             Location/Qualifiers
     source          1..1146
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             1..1146
                     /codon_start=1
                     /product="STARS"
                     /protein_id="AAM27268.1"
                     /db_xref="GI:20530824"
                     /translation="MAPGEKKSSGBGPAKSALRKIRTATLVISLARGWQQWANENSIRQ
                     AQEPTGWLPGGTQDSPQAPKPITPPTSHQKAQSAPKSPPRLPEGHGDGQSSEKAPEVS
                     HIKKKEVSKTVVSKTVERGDVSHLSHRYERDAGVLEPGQPENDIDRILHSHGSPTRR
                     RKCANLVSELIKGWRVMEQEEPTWRSDSVDTEDSGYGGEABERPEQDGVVAVVRIKR
                     PLPSQVNRFTEKLNCKAQQKYSPVGNLKGRWQQWADEHIQSQKLNPFSEEFDYELAMS
                     TRLHKGDEYGRPKEGYKTAERAKRAEEHIYREVMDMCFIICTMARHRRDGKIQVTFG
                     DLFDRYVRISDKVVGIIMRARKHGLVDFEGEMLWQQRDDHVVITLLK"
ORIGIN

Query Match         100.0%;  Score 1146;  DB 9;  Length 1146;
Best Local Similarity  100.0%;  Pred. No. 5.6e-305;
Matches 1145;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATGGCTCCGGGCAAAAAGAAAAGCGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAGATA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTCCGGGCAAAAAGAAAAGCGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAGATA 60

Qy     61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGCCAGCAGTGGGCGAATGAGAAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGCCAGCAGTGGGCGAATGAGAAC 120

Qy    121 AGCATCAGGCAGGCCCAGGAGCCCACAGGCTGGCTGCCCGGGAGGGACCCAGACTCACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGCATCAGGCAGGCCCAGGAGCCCACAGGCTGGCTGCCCGGGAGGGACCCAGACTCACCT 180

Qy    181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACCAGAAAAGCTCAGAGTGCCCCAAAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACCAGAAAAGCTCAGAGTGCCCCAAAG 240

Qy    241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAG 300

Qy    301 GTTTCTCACATCAAAAGAAAGAAGGTGTCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTTCTCACATCAAAAGAAAGAAGGTGTCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360

Qy    361 GGAGGGGACGTGAGCCACCTCAGCCACAGTGATGGGGATGCTGGTGTGCTTGAACCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGAGGGGACGTGAGCCACCTCAGCCACAGTGATGGGGATGCTGGTGTGCTTGAACCT 420

Qy    421 GGGCAGCCAGAGATGACATTGACAGGATCCTCCACAGCCACGCGGAGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGGCAGCCAGAGATGACATTGACAGGATCCTCCACAGCCACGCGGAGG 480

Qy    481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGATGGAGTGAGGAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGATGGAGTGAGGAGGAG 540

Qy    541 GAGCCCACATGGAGGAGTGACACAGCGTAGACAGCAGCAGCAGGCTATGGAGGAGGCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGCCCACATGGAGGAGTGACACAGCGTAGACAGCAGCAGCAGGCTATGGAGGAGGCT 600

Qy    601 GAGGAGAGGGCCCGAGCCAGGCAGGATGCGGCTCGGTGCTGTGTGGTCAGGATCAAGCGCCCCTTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    601 GAGGAGAGGGGCCCGAGCCAGGCAGGATGCGGCTCGGTGCTGTGTGGTCAGGATCAAGCGCCCCTTG 660

Qy    661 CCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC 720

Qy    721 CCAGTGGCAACTTGAAAGGAGGATGGCAGCAGTGGCTGATGAACACATACAATCCAG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CCAGTGGCAACTTGAAAGGAGGATGGCAGCAGTGGCTGATGAACACATACAATCCAG 780

Qy    781 AAGCTCAATCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 840

Qy    841 AAAGGAGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCCAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AAAGGAGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCCAAG 900

Qy    901 CGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG 960

Qy    961 GCTCGCCACAGACGAGGACCAAGATCCAGTTACTTTGGAGATCTCTTTGACAGATAC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCTCGCCACAGACGAGGACCAAGATCCAGTTACTTTGGAGATCTCTTTGACAGATAC 1020

Qy   1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080

Qy   1081 GACTTTGAAGAGAGAATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 GACTTTGAAGAGAGAATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1140

Qy   1141 AAGTGA 1146
          ||||||
Db   1141 AAGTGA 1146
```

```
RESULT 2
AX747674
LOCUS       AX747674                1322 bp    mRNA            linear   PAT 20-JUN-2003
DEFINITION  Sequence 1199 from Patent EP1308459.
ACCESSION   AX747674
VERSION     AX747674.1  GI:32132062
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNA sequences
  JOURNAL   Patent: EP 1308459-A 1199 07-MAY-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1..1322
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match         99.9%;  Score 1144.4;  DB 6;  Length 1322;
Best Local Similarity  99.9%;  Pred. No. 1.5e-304;
Matches 1145;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy      1 ATGGCTCCGGGCAAAAGAAAAGCGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    1141 AAGTGA 1146
           ||||||
Db    1197 AAGTGA 1202


RESULT 3
AK092694
LOCUS       AK092694              1322 bp    mRNA            linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ35375 fis, clone SKMUS2002634.
ACCESSION   AK092694
VERSION     AK092694.1  GI:21751348
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
            Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1322)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..1322
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="SKMUS2002634"
                     /tissue_type="skeletal muscle"
                     /clone_lib="SKMUS2"
                     /note="cloning vector: pME18SFL3"
     CDS             57..1202
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAC03948.1"
                     /db_xref="GI:21751349"
                     /translation="MAPGEKESGEGPAKSALRKIRTATLVISLARGWQQWANENSIRQ
                     AQEPTGWLPGGTQDSPQAPKPITPPTSHQKAQSAPKSPPRLPEGHGDGQSSEKAPEVS
                     HIKKKEVSKTVVSKTYERGGDVSHLSHRYERDAGVLEPGQPENDIDRILHSHGSPTRR
                     RKCANLVSELTKGWRVMEQEEPTWRSDSVDTEDSGYGEAEERPEQDGVVAVVRIKR
                     PLPSQVNRFTEKLNCKAQQKYSPVGNLKGRWCQWADEHIQSQKLNPFSEEFDYELAMS
                     TRLHKGDEGYGRPKEGTKTAERAKRAEEHIYREMDMCFIICTMARHRRDGKIQVTFG
                     DLFDRYVRISDKVVGILMRARKHGLVDFEGEMLWQGRDDHVVITLLK"

ORIGIN
Query Match          99.9%;    Score 1144.4;  DB 9;   Length 1322;
Best Local Similarity  99.9%;    Pred. No. 1.5e-304;
Matches 1145;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy       1 ATGGCTCCGGGCGAAAAGGAAAAGAGCGGGAGGGCCCAGCCAAGAGGCGCCCTCCGGAAGATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      57 ATGGCTCCGGGCGAAAAGGAAAAGAGCGGGAGGGCCCAGCCAAGAGAGGCGCCCTCCGGAAGATA 116
```

```
Qy     61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    117 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAC 176

Qy    121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCTGGGAGGGGACCCAGGACTCACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCTGGGAGGGGACCCAGGACTCACCT 236

Qy    181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCCAAAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 CAAGCTCCTAAACCAATCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCCAAAG 296

Qy    241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCCCTGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    297 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCCCTGAG 356

Qy    301 GTTTCTCACATCAAAAGAAGAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 GTTTCTCACATCAAAAGAAGAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 416

Qy    361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTCTTGTTGAACCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTCTTGTTGAACCT 476

Qy    421 GGGCCAGCAGAGAATGACATTGACAGAATCCTCCACAGCACCGGCTCCCACGCGCGGAGG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 GGGCCAGCAGAGAATGACATTGACAGAATCCTCCACAGCACCGGCTCCCACGCGCGGAGG 536

Qy    481 AGAAAATGTGCCAACCTCGTGTCTCGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    537 AGAAAATGTGCCAACCTCGTGTCTCGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG 596

Qy    541 GAGCCCACAATGGAGGAGTGACACAGCAGGACAGCGGCTATGGAGGAGGAGGCT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GAGCCCACAATGGAGGAGTGACACAGCAGGACAGCGGCTATGGAGGAGGAGGCT 656

Qy    601 GAGGAGGGCCCAGCAGCAGGATGGAGTGCAGGTGCGTGTCAGGATCAAGCGCCCCTTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    657 GAGGAGGGCCCAGCAGCAGGATGGAGTGCAGGTGCGTGTCAGGATCAAGCGCCCCTTG 716

Qy    661 CCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACACAGAAATATAGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    717 CCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACACAGAAATATAGC 776

Qy    721 CCAGTGGGCAACTTGAAAGGGAGAGTGGGCAGCAGTGGGCTGATGAACAATCCCAG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    777 CCAGTGGGCAACTTGAAAGGGAGAGTGGGCAGCAGTGGGCTGATGAACAATCCCAG 836

Qy    781 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTACAC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    837 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTACAC 896

Qy    841 AAAGGAGATGAGGGCTATGGCCGCCCAAAGAGAGGAACCAAAACTGCTGAAAGGGCCAAG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    897 AAAGGAGATGAGGGCTATGGCCGCCCAAAGAGAGGAACCAAAACTGCTGAAAGGGCCAAG 956

Qy    901 CGTGCTGAGGAGCACATCACAGGAAATGAGCACATGTGCTTCATTATCTGCACAATG 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    957 CGTGCTGAGGAGCACATCACAGGAAATGAGCACATGTGCTTCATTATCTGCACAATG 1016

Qy    961 GCTCGCCCACAGCAGAGATGGCAAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGGAGATCTCTTTGACAGATAC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1017 GCTCGCCCACAGCAGACGATGGCAAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGGAGATCTCTTTGACAGATAC 1076

Qy   1021 GTTCGTATTCAGATAAAGTAGTGGCCGCCCAAAGAGAAGGGAACATGGCCAATGACTGGTA 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1077 GTTCGTATTCAGATAAAGTAGTGGCCGCATTCATGCCGTGCCAGGGAAACATGGACACTGGTA 1136

Qy   1081 GACTTTGAAGGAGGAGATGCTATGGCAAGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1137 GACTTTGAAGGAGGAGATGCTATGGCAAGCCGAGATGACCATGTTGTGATTACGCTACTC 1196

Qy   1141 AAGTGA 1146
          ||||||
Db   1197 AAGTGA 1202
```

```
RESULT 4
HSM804735
LOCUS       HSM804735                  2781 bp    mRNA            linear    PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp313H0139 (from clone DKFZp313H0139).
ACCESSION   AL833422
VERSION     AL833422.1  GI:21734063
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2781)
  AUTHORS   Bahr,A., Lauber,J., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from:S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp313H0139) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..2781
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313H0139"
                     /tissue_type="human heart"
                     /clone_lib="313 (synonym: hlcc2). Vector pTriplEx2; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal    2733..2738
     polyA_site      2752
ORIGIN
```

Query Match   99.7%;   Score 1142.8; DB 9;   Length 2781;
Best Local Similarity 99.8%;   Pred. No. 4.3e-304;
Matches 1144;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

```
Qy      1 ATGGCTCCGGGCGAAAAGGAAGACGGGGAGCGGGAGCGGGGAGGGCCCAAGAGCGCCCCTCCGGAAGATA 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     54 ATGGCTCCGGGCGAAAAGGAAGACGGGGAGCGGGAGCGGGGAGGGCCCAAGAGCGCCCCTCCGGAAGATA 113

Qy     61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGGCGGCGAATGAGAAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    114 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGGCGGCGAATGAGAAC 173

Qy    121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCTGGGAGGGGACCCAGGACTCACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    174 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCTGGGAGGGGACCCAGGACTCACCT 233

Qy    181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCCAAAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    234 CAAGCTCCTAAACCAATCACACCCCTACTTCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCCAAAG 293

Qy    241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAACGGTGTCAAAAACCGTGGTGTCCAAAACGGTGATTACGCTACTG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    294 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAACGGTGTCAAAAACCGTGGTGTCCAAAACGGTGATTACGCTACTG 353

Qy    301 GTTTCTCACATCAAAAGAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    354 GTTTCTCACATCAAAAGAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 413

Qy    361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTCTTGAACCT 420
```

COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp686C0314) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES   Location/Qualifiers
     source   1..2832
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686C0314"
                     /tissue_type="human cervix"
                     /clone_lib="686 (synonym: h.icc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal   2735..2740
     polyA_site   2754
ORIGIN

Query Match   99.6%; Score 1141.2; DB 9; Length 2832;
Best Local Similarity 99.7%; Pred. No. 1.2e-303;
Matches 1143; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy    1 ATGGCTCCGGGCGAAAAGGAAGCGGGAGGGCCCAGCCAAGAGCGCCTCCGGAGATA    60
        |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   56 ATGGCTCCGGGCGAAAAGGAAGCGGGAGGGCCCAGCCAAGAGCGCCTCCGGAAGATA  115

Qy   61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCCGAGTGGCAGTGGCGAATGAGAAC  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  116 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCCGAGTGGCAGTGGCGAATGAGAAC  175

Qy  121 AGCATCAGGCAGGCCCAGGAGCTCACAGGCTCGCTGCCGGAGGGACCCAGGACTCACCT  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  176 AGCATCAGGCAGGCCCAGGAGCTCACAGGCTCGCTGCCGGAGGGACCCAGGACTCACCT  235

Qy  181 CAAGCTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAG  240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  236 CAAGCTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAG  295

Qy  241 TGCCACCCGCTGCCAGACGACATGGACATGGACAAGGTGTCCAAAAACGGTGGTCAGACTTACGAGAGA  300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  296 TGCCACCCGCTGCCAGACGACATGGACATGGACAAGGTGTCCAAAAACGGTGGTCAGACTTACGAGAGA  355

Qy  301 GTTTCTCACATCAAAAGAAAGAGGTGTCCAAAAACGGTGGTCAGACTTATGAGAGA  360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  356 GTTTCTCACATCAAAAGAAAGAGGTGTCCAAAAACGGTGGTCAGACTTATGAGAGA  415

Qy  361 GGAGGGGACGTGAGCCACCTCAGCCAGTGGCCACCCAGGATGCTGGTGTGCTTGAACCT  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  416 GGAGGGGACGTGAGCCACCTCAGCCAGTGGCCACCCAGGATGCTGGTGTGCTTGAACCT  475

Qy  421 GGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCAAGGGCTGAGAGTGATGAGGAG  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  476 GGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCAAGGGCTGAGAGTGATGAGGAG  535

Qy  481 AGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG  540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  536 AGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG  595

Qy  541 GAGCCCACATGGAGGAGGCGTAGACAGAGGACACAGCCGCTATGGAGGAGGAGGCT  600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  596 GAGCCCACATGGAGGAGGCGTAGACAGAGGACACAGCCGCTATGGAGGAGGAGGCT  655

Qy  601 GAGGAGAGGCCCGAGCAGGCCCGAGCAGGTGCGTGTGGCTCAGGATCCAAGCGCCCCTTG  660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  656 GAGGAGAGGCCCGAGCAGGCCCGAGCAGGTGCGTGTGGCTCAGGATCCAAGCGCCCCTTG  715

Qy  661 CCCTCCCAGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  720
```

```
Db  414 ...GGGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGAGGGATGCTGGTGTGCTTGAACCT  473
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  421 GGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGGCTGAGAGTGATGAGAGG  480

Db  474 GGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGGCTCCCACACGCGGAGG  533
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  481 AGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG  540

Db  534 AGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG  593
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  541 GAGCCCACATGGAGGAGGCGTAGACAGAGGACACAGCCGCTATGGAGGAGGAGGCT  600

Db  594 GAGCCCACATGGAGGAGGCGTAGACAGAGGACACAGCCGCTATGGAGGAGGAGGCT  653
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  601 GAGGAGAGGCCCGAGCAGGCCCGAGCAGGTGCGTGTGGCTCAGGATCCAAGCGCCCCTTG  660

Db  654 GAGGAGAGGCCCGAGCAGGCCCGAGCAGGTGCGTGTGGCTCAGGATCCAAGCGCCCCTTG  713
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  661 CCCTCCCAGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  720

Db  714 CCCTCCCAGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  773
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  721 CCAGTGGGCAACTTGAAAGGGAGATGGGCGAGTGGCCAGTGCAGTGGCCGCAGTGGGCTGCAGTGAACACATACAATCCCAG  780

Db  774 CCAGTGGGCAACTTGAAAGGGAGATGGGCGAGTGGCCAGTGCAGTGGCCGCAGTGGGCTGCAGTGAACACATACAATCCCAG  833

Qy  781 AAGCTCAATCCTTTCAGTGAAGAGTTGATTACCAGCTGGCCATGTCCACCCGCTACAC  840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  834 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACCAGCTGGCCATGTCCACCCGCTACAC  893

Qy  841 AAAGGAGATGAGGCTATGGCCGCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  900
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  894 AAAGGAGATGAGGGCTATGGCCGTCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  953

Qy  901 CGTGCTGAGGAGCACATCACAGGGACACATCTCACAGGGAAATGATGCTTCATTATCTGCACAATG  960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  954 CGTGCTGAGGAGCACATCACAGGGACACATCTCACAGGGAAATGATGCTTCATTATCTGCACAATG  1013

Qy  961 GCTCGCCACAGACGAGATGGCAAGATCCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATAC  1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1014 GCTCGCCACAGACGAGATGGCAAGATGCAAGATCCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATAC  1073

Qy 1021 GTTCGTATTTCAGATAAAGTAGTAGTGGGCATTCTCATGGCGTGGCAAGGCGGAAACATGGACTGGTA  1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1074 GTTCGTATTTCAGATAAAGTAGTAGTGGGCATTCTCATGGCGTGGCAAGGCGGAAACATGGACTGGTA  1133

Qy 1081 GACTTTGAAGGAGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGATTACGGCTACTC  1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1134 GACTTTGAAGGAGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTTGATTACGGCTACTC  1193

Qy 1141 AAGTGA  1146
        ||||||
Db 1194 AAGTGA  1199
```

RESULT 5
HSM803459

LOCUS   HSM803459   2832 bp   mRNA   linear   PRI 13-MAY-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp686C0314 (from clone DKFZp686C0314).
ACCESSION   AL832152
VERSION   AL832152.1 GI:21732696
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2832)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY

```
AQEPAGWLPGATHDVPNAPKEAGPYQHAPKTLSFKPDRDGEGQHSEEATEVSHIKRKE
VTRTVVSKAYERGGDVNYLSHRYENDGGVSEAIQPENDIDRILLSHDSPTRRKCTNL
VSELTKGWKVMEQEEPTWKSDSVDTEDSGYGGDMERPEQDAAPVAPARIKRPLHSQA
NRYSEPLNCKAHRKYSQVDNLKGRWQQWADEHVQSQKLNPFSDEFPYDLAMSTRLHKG
DEGYGRPKEGSKTAERAKRAEEHIYREIMELCFVIRTMARHRRDGKIQVTFGELFPRY
VRISDKVVGILMEARKGHLVHFEGEMLWQGRDHVITLVE"

ORIGIN

Query Match        69.0%;    Score 791.2;  DB 10;   Length 1128;
Best Local Similarity  81.6%;   Pred. No. 5e-207;
Matches  935;  Conservative  0;  Mismatches 193;  Indels 18;  Gaps  1;

Qy     1 ATGGCTCCGGCGAAAGAAGGAAAGCGGGAGGGCCAGCCAAGCCAGCCCTCCGGAAGATA   60
         ||||||||| || |||| |||| |||   |||||  || || || || ||||||||| |
Db     1 ATGGCTCCAGGAGAAAGGAAAGGAAAGGGAGGGCGGCGGCCAAGAGTGCCCTCCGGAAGTC   60

Qy    61 CGCACAGCCACCTGGTCATCAGCTTGGCGCAGCTTGGCGCAGTGGCGAATGAGAGAC  120
         |||||||| ||| ||| ||| | |||| |||||||||||||||||||||||||||||
Db    61 CGCACAGCAACCCTGGTTATCAATTGGCCCAGCTTGGCGCAGTGGCGAATGAGAGAC  120

Qy   121 AGATCAGGCAGGCCCAGAGCCTCACAGGCTGGCTGCCGGGGAGGGACCCAGGACTCACCT  180
         || | |||||| || || | |||  |  |||| || ||||| ||||| || |  ||||||
Db   121 AGTACCAAACAGGCCCAGAGCCTGGCAGGCTGGCTGCCGGGGAGGGAGCAACTCATGACGTACCT  180

Qy   181 CAAGTCCTAAACAATCACACCCCCTACTTCACACGAAAAGTCAGATGCCCCAAAG  240
                                   |||||||||||||||| || ||||||||
Db   181 AACGGTCCTAAAGAAGAAGCCCGTCC---------------TTACCAGCATGCCCCAAA  222

Qy   241 TCGCCACCCGCCTGCCGAAGGACATGGAAGGACACGGAGACGGAGGGACAACACTCAGAGAAGAAGCCACCGAG  300
         ||| ||  | ||| |||||||||||  | ||||||||||||| || |||| |||| ||||||||||||||||||
Db   223 ACTCTGTCTCCAAAGCCAGATCGAGACGGAGGGACAACACTCAGAGAAGAAGCCACCGAG  282

Qy   301 GTTTCTCAATCAAAAAGAAAGAGAGTGTCCAAAACGGTTGGTCAAGACTTACGAGAGA  360
         ||||||||| |||||||||||||||||||| |||||||||||||||| |||||||||||
Db   283 GTCTCCCACATTAAAGAAAGAGAAGTGACCAGAACGGTTGTCAGCAAGACGGTTGTCAGCAAGGCTTATGAGAGG  342

Qy   361 GGAGGGACGTGAGCCACCTCAGCCACAGTGGTACCAGCAGTGGGATGTGAGGAGATGCTTGAACCT  420
         ||||| |||||||||||| || |||| ||||||| ||| |||| || || ||| ||| |||| |||
Db   343 GGAGGAGAGATGCTGAACTACCTGAGCCACAGTGATGATGATGATGGAGGAGAAGGCT  402

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGCCGAGG  480
         ||||| ||||||| ||||| ||| || |||  | | ||||| |||| ||
Db   403 ATTCAGCCAGAGAATGACATTGACAGAATCCTTAGTCACGACTCGCCAACACGAGA  462

Qy   481 AGAAAAATGTGCCAACCTGGTGTCTGAGCTAACAGAGGTGATGGGACCAGGAG  540
         |||||||||  |||||||||| || |||  || || |||||||||| |||||
Db   463 AGAAAAATGACCACCCTGGTGTCTGAGCTGACCAAAGGCTGGAAAGTGATGAAACAGGAA  522

Qy   541 GAGCCCACATGGAGGAGTGACAGCGCTAGAACACAGAGACACACACAGTGGGCTATGGAGAGAGGCT  600
         || |||||||||| ||| ||| || | ||| | ||| |||||| ||||||||| ||||||||||||
Db   523 GAGCCCAGTGGAAGAGTGACAGCGCAGAGTGACAGAGGGGATATG  582

Qy   601 GAGGAGGCCGAGCAAGCAGCAGATGGAGTGTCAGGTGGCTCGGATCAAGCAAGCGCCCCCTTG  660
         |||||||| ||| ||| |||| | ||||||||||| |||| ||| |||| ||||| ||| |||
Db   583 GAGGAGGCCGAGCAAGAATGCAAGAATGCTACGGCGCCTCGTGGCCTCGGATCAAACGCCCCCTG  642

Qy   661 CCCTCCCAGTGTAAACAGATTTACAGAGAAACTCAACTGCAAGGCCCAACAGAAATAGC  720
         ||| ||||||| | | ||| |||| || | ||||||||||| ||||| || ||||||||
Db   643 CACTCCCAGGCAAACAGGTACTCTGAGCCACTCAACTGTAAGGCCCATCGTAAGGCCAAATACAGC  702

Qy   721 CCAGTGGGCAACTTGAAAGGAGATGGCGCAGCAGTGGCTGATGAACACATCAATCCCAG  780
         ||||| ||| ||| |||| |||| || ||||| ||||| ||  |||||| |||| ||||
Db   703 CAAGTTGACAACTTGAAAGGAGGTGGCAGCAGAGTTTGATTACGACCCGCTACAC  762

Qy   781 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTACAC  840
         ||||| |||| ||| |||| |||| |||||||||||| ||| |||||| || |||||||
Db   763 AAGCTTCAATCCTTCAGTGACGAATTTGACTATGACCTAGCCATGTCCACTCGGCTCCAC  822

Qy   841 AAAGGAGATGAGGGGCTATGGCCGCGCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  900
         ||| ||||||||| ||||||||||||||| ||||||| ||| || ||||||||| ||||||
Db   823 AAGGGAGACGGGGGCTATGGCCGCGCCCCCAAAGAGAGAGAAGCAAGAGCATGGGCCAAG  882

Qy   901 CGTGCTGAGGAGCACATCACAGGGAAATGATGACACGTCGCTTCATTATCTGCACAATG  960
         ||||| ||||||||||||||||||||||||||||| || |||||||||||||||||||
Db   901 CGTGCTGAGGAGCACATCACAGGGAAATGATGACACGTCGCTTCATTATCTGCACAATG  960

Db   716 CCCTCCCAGTGTAAACAGATTTACAGAGAAACTCAACTGCAAGGCCCAACAGAAATATAGC  775
Qy   721 CCAGTGGGCAACTTGAAAGGAGATGGCGCAGCAGTGGCTGATGAACACATCAATCCCAG  780
Db   776 CCAGTGGGCAACTTGAAAGGAGATGGCGCAGCAGTGGCTGATGAACACATCAATCCCAG  835
Qy   781 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTACAC  840
Db   836 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTACAC  895
Qy   841 AAAGGAGATGAGGGGCTATGGCCGCGCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  900
Db   896 AAAGGAGATGAGGGGCTATGGCCGCGCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  955
Qy   901 CGTGCTGAGGAGCACATCACAGGGAAATGATGACACGTCGCTTCATTATCTGCACAATG  960
Db   956 CGTGCTGAGGAGCACACATCACAGGGAAATGATGACACGTCGCTTCATTATCTGCACAATG 1015
Qy   961 GCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTCTTTGAGATCTCTTTGACAGATAC 1020
Db  1016 GCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTCTTTGAGATCTCTTTGACAGATAC 1075
Qy  1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080
Db  1076 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1135
Qy  1081 GACTTTGAAGGAGGAGATGCTATGGCCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
Db  1136 GACTTTGAAGGAGGAGATGCTATGGCCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1195
Qy  1141 AAGTGA 1146
Db  1196 AAGTGA 1201
```

RESULT 6
AF504061
LOCUS   AF504061   1128 bp   mRNA   linear   ROD 01-JUL-2002
DEFINITION Mus musculus striated muscle activator of Rho-dependent signaling
            mRNA, complete cds.
ACCESSION   AF504061
VERSION   AF504061.1 GI:20805880
KEYWORDS   .
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1128)
AUTHORS   Arai,A., Spencer,J.A. and Olson,E.N.
TITLE   STARS, a Striated Muscle Activator of Rho Signaling and Serum
            Response Factor-dependent Transcription
JOURNAL   J. Biol. Chem. 277 (27), 24453-24459 (2002)
MEDLINE   22086201
PUBMED   11983702
REFERENCE   2 (bases 1 to 1128)
AUTHORS   Arai,A., Spencer,J.A. and Olson,E.N.
TITLE   Direct Submission
JOURNAL   Submitted (19-APR-2002) Molecular Biology, University of Texas
            Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
            75390-9148, USA
FEATURES   Location/Qualifiers
     source   1..1128
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS   1..1128
                     /note="STARS"
                     /codon_start=1
                     /product="striated muscle activator of Rho-dependent
                     signaling"
                     /protein_id="AAM28877.1"
                     /db_xref="GI:20805881"
                     /translation="MAPGERERERGPAKSALRKVRTATLVINLARGWQQWANENSTKQ

```
Db   883 CGAGCGGAAGAGCACATCTATCGGGAAATTATGAACTGTGCTTTGTTGTTATCCGCACAATG 942
             ||  ||||   |||||||||||||||||  ||||| |||||||||||||||||  |||||||||
Qy   961 GCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTCGGAGATCTCTTTGACAGATAC 1020
             |||||||||||||||  ||| ||||||||||||||||||||||||||| |||||||  ||  |
Db   943 GCTCGCCACAGACGACGACGAGATGCAAGATCCAGGTTACTTTCGGAGAACTCTTTGATCGCTAT 1002
Qy  1021 GTTCGTATTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080
             ||||| ||| | |||||||| |||||||| |||||||||||||||||| ||||||||| 
Db  1003 GTTCGCATTTCTGATAAAGTCGTGGGCATCCTCATGCGTGCCAGGAAACACGGACTGGTG 1062
Qy  1081 GACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGATGACCACCATGTTGTGATTACGCTACTC 1140
             |||||||||||| |||||||||||||||||||||| ||||||||||||||||||||| | | 
Db  1063 CACTTTGAAGGAGGAGATGCTATGGCAAGGCCGAGACGACCACCATGTTGTGATTACTCTCGTT 1122
Qy  1141 AAGTGA 1146
               | ||
Db  1123 GAGTAA 1128
```

RESULT 7

AF336113

```
LOCUS       AF336113              1285 bp    mRNA    linear   ROD 13-AUG-2002
DEFINITION  Rattus norvegicus MS1 mRNA, complete cds.
ACCESSION   AF336113
VERSION     AF336113.1  GI:22212591
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1285)
  AUTHORS   Mahadeva,H., Brooks,G., Lodwick,D., Chong,N.W. and Samani,N.J.
  TITLE     MS1, a novel stress-responsive, muscle-specific gene that is
            up-regulated in the early stages of pressure overload-induced left
            ventricular hypertrophy
  JOURNAL   FEBS Letters (2002) In press
REFERENCE   2  (bases 1 to 1285)
  AUTHORS   Mahadeva,H. and Samani,N.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2001) Department of Cardiology, Leicester
            University, Groby Road, Leicester LE3 9QP, England
FEATURES             Location/Qualifiers
     source          1..1285
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="WKY"
                     /db_xref="taxon:10116"
     CDS             54..1181
                     /codon_start=1
                     /product="MS1"
                     /protein_id="AAM94370.1"
                     /db_xref="GI:22212592"
                     /translation="MAPGSTVREAGPAKSALQKVRRATLVINLARGWQQWANENSTRQ
                     AQEPAGWLPGATQDLPHTPKEPGPRQHAPKPPSPKPDGDRREGRGSEBATEVSHIKRKE
                     VTRTVVSKAYERGGDVNYLSHRYEHDGGVSEAVQPDNDIDRILLSHDSPTRRKCTNL
                     VSKLTKGWKVMEOEPEKWKSDSIDTEDSGYGGDMEERPEQDVAQVAAARIKRPLHSQA
                     NRYSETLNCKAHRKYSQVDNLKGRWQQWADEHIQSQKLNPFSDEFDYDLAMSTRLHKG
                     DEGYGRPKEGSKTAERAKRAEEHIYREIMELCPVIRTMARHRRDGKIQVTFGELFPDRY
                     VRISDKVVGILMRARKGLVHFEGEMLWQGKDHVVITLLE"
ORIGIN
```

Query Match 68.5%; Score 784.8; DB 10; Length 1285;
Best Local Similarity 81.2%; Pred. No. 2.9e-205;
Matches 931; Conservative 0; Mismatches 197; Indels 18; Gaps 1;

```
Qy     1 ATGGCTCCGGGCGAAAAGGAAAGCCGGGAGAGCGGCGGCGGGAAGAGCGCCCTCCGGAAGATA 60
             |||||||| ||| |||||  |||| ||||||||||||  |||   ||||||  |||||||  
Db    54 ATGGCTCCTGGGAGAAAACTGTAAGGGAGGCGGCGGGCGGGCCGCCCCGGCGGAGAGTGCCCCTCCAGAAGGTC 113
Qy    61 CGCAGCAGCCACCCTGGTCATCAGCAGTTGGCCCGAGTTGGCCAGTGGGCCGGAGAATGAGAAC 120
             |||||  ||||||  |||| |||||||  ||  ||   ||||||||||||  |||||||||||
Db   114 CGCAGCAGCCACCCTGGTCATCAGCAGTTGGGCCCAGTTGGCCAGTGGGCCGGAGAATGAGAAC 173
Qy   121 AGCATCAGGCAGGCCCAGGAGCTGCTACAGGCTGGCGGGAGGGGGACCCAGGACTCACCT 180
             || |||||| ||||||||||||||| ||||||| || |||||  |  |||| |||  ||||
Db   174 AGTACCAGACAGCCCAGGAGCTGCAGGCTGCTGCCAGGAGCAACTCAAGACCTACCT 233
Qy   181 CAAGCTCCTAAACCAATCACACCCCTACTTCACCAGAAAGCTCAGAGTGCCCCAAAG 240
             |||  |||||||   | ||  || | | ||           |||  |||||||||
Db   234 CACACTCCTAAAGAACCGGGTCCTC-----------GCCAGCATGCGCCCCAAA 275
Qy   241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGTCAGAGAAAGCCCTGAG 300
             || ||  |||||  ||||| ||  || |||||||| || ||||| |||||||| |||
Db   276 CCTCCGTCTCCAAAGCCAGATGAGACAGAGGGACGAGCTCGAGGAAGCCACGGAG 335
Qy   301 GTTTCTCACATCAAAAGAAAAGAGGTGTCAAAACGGTGTCAAGACTTACGAGAGA 360
             || ||| || ||||||||||||| |||| | ||||| ||||  || | ||||||
Db   336 GTCTCCCATATCAAAAGAAAAGAGGTGACCAGACAGTTGTCAGTCAGTGATGAGAGA 395
Qy   361 GGAGGGGACGTGAGCCACCTCAGCCAGGTACGACAGGTACGAGAGGGATGCTGGTGGCGGG 420
             |||||  |||||  | |||| |||||| |||||| |||||||| ||  ||||| | |||||
Db   396 GGAGGAGTGTGAACTACCTGAGCCACAGGTACGAGAGGTACGAGCATGATGCGGGCGTGCTGAAGCG 455
Qy   421 GGGCAGCCAGAAGAATGACATTGACAGAATCCTCCACGACAGTCCCCACGCCGGAGG 480
             | ||||||||||||||||||||||||||||||||| | |   |  | ||| ||||
Db   456 GTCCAGCCAGAAGAATGACATTGACAGAATCCTCCTCAGTCACGACTCGCCAACGCGGAGA 515
Qy   481 AGAAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAAAGTGATGGAGCAGAGG 540
             ||||| ||| | ||| ||||| |||| ||| ||||| |||||||||||||||| ||| ||
Db   516 AGAAAATGACCACCAACTTGGTATCTAAGCTGACCAAAGGCTGGAAAGTGATGGAACAGGAG 575
Qy   541 GAGCCCACATGGAGAGTGACAGCTAGACACAGAGACACAGAGGCAGTGGCTACGGAGGGGCATG 600
             |||||||  ||| ||||||||| | ||| |||||| |||||||| |||| ||||||||  || 
Db   576 GAGCCCAGTGAAGAGTGACAGCATAGACACAGAGACACAGAGGCAGTGGCTACGGAGGGAGAGCT 635
Qy   601 GAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGTTCAGGATCAAGCGCCCTTG 660
             ||||| ||||||||||| |||| ||||||||||||| ||| |||| | |||||
Db   636 GAGGAGAGGCCCGAGCAAGTAGTGCGCAGGTGGCTGCGCCAGGATTAAACGCCCCTTG 695
Qy   661 CCCTCCCAGGTAAACAGATTTACAGAGAAACTGCAAAGCCCAACAGAGAAATATAGC 720
             |||  ||||||||||||||| ||||| ||||| | || |||| || |||||| | |
Db   696 CACTCCCAGGCAAACAGATACTCAGAGACACTCAACTGTAAGGCCCATCGGAAATACAGC 755
Qy   721 CCAGTGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACATCACCAG 780
             |||||| |||| |||||||| |  ||||||||||| ||||| |||||| |||
Db   756 CAAGTGGACAACTTGAAAGGCCGATGGCAGCAGTGGGCCGATGAACATATACAGTCACAG 815
Qy   781 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGTCCACCCGCCTACAC 840
             ||||||| || | ||||| |||||||| |||| ||| |||||| ||| ||
Db   816 AAGCTCAACCCCTTCAGTGAAGATTTCGACTATGACCTAGCCATGTCCACTCGACTCCAC 875
Qy   841 AAAGGAGATGAGGGCTATGGCCGCCCAAAACTGCTGAAAGGGCCAAG 900
             ||| |||| ||| || || || || |||| | |   ||| |||||||
Db   876 AAGGGAGACGAGGGATATGGCCGCCCAAAAGAGGGAAGCAAGACAGCAGAGGCAAG 935
Qy   901 CGTGCTGAGGACGCATCTCACAGGGAAATGATGCTTCATTATCTGCACAATG 960
             ||| ||||||||||||| | |||  ||||| |  | | || | | || |||||
Db   936 AGAGCGGAGGAGCACATCTATCGGGAAATTGTGAATTGTGCTTTGTTATCCCACAATG 995
Qy   961 GCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTCGGAGATCTCTTTGACAGATAC 1020
             |||||||||||||| ||||||||| ||||||||| |||||||||||| |||||| ||||
Db   996 GCTCGCCATAGACGAGATGGCAAGATCCAGGTTACTTTCGGAGAACTCTTTGATAGATAT 1055
Qy  1021 GTTCGTATTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080
             ||||| ||| | |||||||| |||||||| |||||||||||||||||| ||||||||| 
Db  1056 GTTCGAATTCAGATAAAGTCGTGGGCATCCTCATGCGTGCCAGGAAAGACGGACTGGTG 1115
Qy  1081 GACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGATGACCACCATGTTGTGATTACGCTACTC 1140
             |||||||||||| |||||||||||||||||||||| ||||||||||||||||||||| | | 
Db  1116 CACTTTGAAGGGGAGATGCTATGGCAAGGCCGAGACGACCATGTTGTGATTACTCTCCTT 1175
Qy  1141 AAGTGA 1146
               | ||
Db  1176 GAGTAA 1181
```

```
RESULT 8
AP000856/c
LOCUS       AP000856              152130 bp    DNA     linear   PRI 01-MAR-2001
DEFINITION  Homo sapiens genomic DNA, chromosome 8q23, clone:KB1539E1.
ACCESSION   AP000856
VERSION     AP000856.3  GI:13358956
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
AUTHORS     Shimizu,N. and Asakawa,S.
TITLE       Homo sapiens DNA chromosome 8 SEQUENCE
JOURNAL     Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 152130)
AUTHORS     Shimizu,N. and Asakawa,S.
TITLE       Direct Submission
JOURNAL     Submitted (13-DEC-1999) Nobuyoshi Shimizu, Keio University, School
            of Medicine, Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo
            160-8582, Japan (E-mail:nshimizu@dmb.med.keio.ac.jp,
            Tel:81-3-3351-2370, Fax:81-3-3351-2370)
COMMENT     On Mar 16, 2001 this sequence version replaced gi:8096485.
FEATURES             Location/Qualifiers
     source          1..152130
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="KB1539E1"
                     /cell_line="FLEB 14 - 14"
                     /clone_lib="Keio BAC library"
     repeat_region   1603..1809
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   1819..2094
                     /evidence=not_experimental
                     /rpt_family="AluSg1"
     repeat_region   2098..2136
                     /evidence=not_experimental
                     /rpt_family="(CAAA)n"
     repeat_region   complement(3276..3445)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   3528..3670
                     /evidence=not_experimental
                     /rpt_family="L1MC5"
     repeat_region   3737..3827
                     /evidence=not_experimental
                     /rpt_family="MER5A"
     repeat_region   3888..4246
                     /evidence=not_experimental
                     /rpt_family="L1MC5"
     repeat_region   complement(4307..4612)
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   complement(6251..6622)
                     /evidence=not_experimental
                     /rpt_family="LTR37A"
     repeat_region   complement(6627..6929)
                     /evidence=not_experimental
                     /rpt_family="AluY"
     repeat_region   complement(7278..7453)
                     /evidence=not_experimental
                     /rpt_family="MER5A"
     repeat_region   7484..7774
                     /evidence=not_experimental
                     /rpt_family="AluSg"
     repeat_region   complement(7792..7991)
                     /evidence=not_experimental

     repeat_region   complement(8094..8190)
                     /evidence=not_experimental
                     /rpt_family="LTR54"
     repeat_region   complement(8417..8474)
                     /evidence=not_experimental
                     /rpt_family="MER10"
     repeat_region   9119..9260
                     /evidence=not_experimental
                     /rpt_family="(TA)n"
     repeat_region   complement(10809..10935)
                     /evidence=not_experimental
                     /rpt_family="FLAM_A"
     repeat_region   complement(11155..11575)
                     /evidence=not_experimental
                     /rpt_family="MLT1J1"
     repeat_region   complement(12042..12257)
                     /evidence=not_experimental
                     /rpt_family="MER58A"
     repeat_region   12298..12319
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   complement(12382..12476)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   12605..12686
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(13249..13444)
                     /evidence=not_experimental
                     /rpt_family="MLT1I"
     repeat_region   13704..13811
                     /evidence=not_experimental
                     /rpt_family="MER5A"
     repeat_region   complement(13841..13951)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   complement(14971..15272)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(16581..16677)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   17788..18078
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   18079..18098
                     /evidence=not_experimental
                     /rpt_family="(CAA)n"
     repeat_region   18556..18577
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   complement(18578..18860)
                     /evidence=not_experimental
                     /rpt_family="AluY"
     repeat_region   18861..18882
                     /evidence=not_experimental
                     /rpt_family="AT_rich"
     repeat_region   complement(19014..19214)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(19652..20150)
                     /evidence=not_experimental
                     /rpt_family="L1MB5"
     repeat_region   complement(20157..20921)
                     /evidence=not_experimental
                     /rpt_family="L1M4"
     repeat_region   20936..21000
                     /evidence=not_experimental
                     /rpt_family="MER91B"
     repeat_region   21203..23651
                     /evidence=not_experimental
                     /rpt_family="L1M4"
```

```
repeat_region   23654..24174
                /evidence=not_experimental
                /rpt_family="L1MB8"
repeat_region   complement(24671..24701)
                /evidence=not_experimental
                /rpt_family="L2"
repeat_region   24911..24932
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   complement(25053..25411)
                /evidence=not_experimental
                /rpt_family="MLT1A1"
repeat_region   25476..25577
                /evidence=not_experimental
                /rpt_family="MER5A"
repeat_region   complement(25502..25607)
                /evidence=not_experimental
                /rpt_family="MER5A"
repeat_region   25585..25705
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   complement(28001..28123)
                /evidence=not_experimental
                /rpt_family="FLAM_C"
repeat_region   28582..28684
                /evidence=not_experimental
                /rpt_family="FLAM_A"
repeat_region   28892..28972
                /evidence=not_experimental
                /rpt_family="FLAM_A"
repeat_region   29596..29893
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   29894..30069
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   31207..31232
                /evidence=not_experimental
                /rpt_family="AluSg/x"
repeat_region   complement(31446..31520)
                /evidence=not_experimental
                /rpt_family="AT_rich"
repeat_region   31705..31746
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   31775..31993
                /evidence=not_experimental
                /rpt_family="TATGln"
repeat_region   complement(32042..32122)
                /evidence=not_experimental
                /rpt_family="MER46A"
repeat_region   32464..32763
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(33126..33290)
                /evidence=not_experimental
                /rpt_family="MER1B"
repeat_region   complement(33707..34067)
                /evidence=not_experimental
                /rpt_family="MER5A"
```

```
Query Match  58.2%;  Score 667.4;  DB 9;  Length 152130;
Best Local Similarity 99.9%;  Pred. No. 7.7e-173;
Matches 668;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0;

Qy       1 ATGGCTCCGGGCGAAAAGGAAAAGCGGGAAGGCGGGAGGGCCCTTCCGGAAGAGATA 60
           ||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||
Db  139073 ATGGCTCCGGGCGAAAAGGAAAAGCGGGAAGGCGGGAGGGCCCCTCCGGGAAGAGATA 139014

Qy      61 CGCACAGCCACCCTGCTCATCAGCTTGGCCCGAGGTTGGCAGTGGGCGAATGAGAAC 120
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  139013 CGCACAGCCACCCTGCTCATCAGCTTGGCCCCAGGGTTGGCAGTGGGCGAATGAGAAC 138954

Qy     121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTTGGCTCGCGGGAGGGACCCCAGGACTTCACCT 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  138953 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTTGGCTCGCGGGAGGGGACCCCAGGACTCACCT 138894
Qy     181                                                                 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138893 CAAGCTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAG 138834
Qy     241                                                                 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138833 TCGCCACCCGCCTGCGCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAAGCCCTGAG 138774
Qy     301                                                                 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138773 GTTTCTCACATCAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 138714
Qy     361                                                                 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138713 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTGGTTGAACCT 138654
Qy     421                                                                 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138653 GGGCAGCCAGAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGCGGAGG 138594
Qy     481                                                                 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138593 AGAAAATGTGCCAACCTGGTGTCGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG 138534
Qy     541                                                                 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138533 GAGCCCACATGAGGGAGTGACAGCCGTAGACACAGAGACAGCCGGCTATGGAGAGGAGGCT 138474
Qy     601                                                                 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  138473 GAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGCGTGTGGTCAGGATCAAGCGCCCCTTG 138414
Qy     661 CCCTCCCAG 669
           |||||||||
Db  138413 CCCTCCCAG 138405
```

```
RESULT 9
AF303733/c
LOCUS          AF303733          153147 bp    DNA    linear    HTG 26-JUL-2002
DEFINITION     Homo sapiens chromosome 8 clone RP11-79F7 map 8q23, WORKING DRAFT
               SEQUENCE, 10 unordered pieces.
ACCESSION      AF303733
VERSION        AF303733.2  GI:14389401
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_CANCELLED.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 153147)
AUTHORS        Reichwald,K., Baumgart,C., Blechschmidt,K., Dette,M., Jahn,N.,
               Lehmann,R., Menzel,U., Polley,A., Schilhabel,M.B., Schudy,A.,
               Siddiqui,R., Taudien,S., Wen,G., Rosenthal,A. and Platzer,M.
TITLE          Chromosome 8 genomic sequence
JOURNAL        Unpublished
REFERENCE      2  (bases 1 to 153147)
AUTHORS        Genome Sequencing Center Jena.
TITLE          Direct Submission
JOURNAL        Submitted (05-SEP-2000) Genome Analysis, Institute of Molecular
               Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT        On Jun 13, 2001 this sequence version replaced gi:10442740.
               -------------- Genome Center
                  Center: Institute of Molecular Biotechnoloy
                  Center code: IMB
                  Web site: http://genome.imb-jena.de/
                  Contact: gscj-submit@genome.imb-jena.de
               -------------- Project Information
                  Center project name: H552
                  Center clone name: RP11-79F7
               -------------- Summary Statistics
                  Sequencing vector: M13; 100% of reads
```

```
Qy    121 AGCATCAGGCAGGCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGAGGACCCAGGACTCACCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98366 AGCATCAGGCAGGCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGAGGACCCAGGACTCACCT 98307

Qy    181 CAAGCTCCTAAACCAATCACACCCCTACTTCACCAGAAGCTCACCAGAAGCTCAGAGTGCCCAAAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98306 CAAGCTCCTAAACCAATCACACCCCTACTTCACCAGAAGCTCACCAGAAGCTCAGAGTGCCCAAAG 98247

Qy    241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98246 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAG 98187

Qy    301 GTTTCTCACATCAAAAGAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98186 GTTTCTCACATCAAAAGAAAGAAAGAGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 98127

Qy    361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTAGCAGAGGGATGCTGGTGCTTGAACCT 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98126 GGAGGGGACGTGAGCCACCTCAGCCACAGGTAGCAGAGGGATGCTGGTGCTTGAACCT 98067

Qy    421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGCGGAGG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98066 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGCGGAGG 98007

Qy    481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGTGAGGGCTGGAGGAGGAG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  98006 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGTGAGGGCTGGAGGAGGAG 97947

Qy    541 GAGCCCACATGGAGAGTGACACGGCTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97946 GAGCCCACATGGAGAGTGACACGGCTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCT 97887

Qy    601 GAGGAGGGCCGAGCAGGAGTGCAGGTGCTGTTGCTCAGGATCAAGCGCCCCTTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  97886 GAGGAGGGCCGAGCAGGAGTGCAGGTGCTGTTGCTCAGGATCAAGCGCCCCTTG 97827

Qy    661 CCCTCCCAG 669
          |||||||||
Db  97826 CCCTCCCAG 97818
```

RESULT 10
AC023344/c
```
LOCUS       AC023344          186751 bp    DNA     linear   PRI 09-MAY-2001
DEFINITION  Homo sapiens BAC clone RP11-395G23 from 8, complete sequence.
ACCESSION   AC023344
VERSION     AC023344.4   GI:11120953
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 186751)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 186751)
  AUTHORS   Fryman,J., Drone,K., Haakenson,W. and Bartniczak,K.
  TITLE     The sequence of Homo sapiens BAC clone RP11-395G23
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 186751)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 186751)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
```

```
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 145861 bases at least Q40
            Consensus quality: 148833 bases at least Q30
            Consensus quality: 150578 bases at least Q20
            Quality coverage: 4.94 x in Q20 bases; sum-of-contigs

            ----------------------------------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality 10.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
            ----------------------------------------

            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1   1899:  contig of 1899 bp in length
            *     1900   1999:  gap of unknown length
            *     2000   5899:  contig of 3900 bp in length
            *     5900   5999:  gap of unknown length
            *     6000  15182:  contig of 9183 bp in length
            *    15183  15282:  gap of unknown length
            *    15283  30210:  contig of 14928 bp in length
            *    30211  30310:  gap of unknown length
            *    30311  44943:  contig of 14633 bp in length
            *    44944  45043:  gap of unknown length
            *    45044  71733:  contig of 26690 bp in length
            *    71734  71833:  gap of unknown length
            *    71834 105674:  contig of 33841 bp in length
            *   105675 105774:  gap of unknown length
            *   105775 128636:  contig of 22862 bp in length
            *   128637 128736:  gap of unknown length
            *   128737 148807:  contig of 20071 bp in length
            *   148808 148907:  gap of unknown length
            *   148908 153147:  contig of 4240 bp in length.

FEATURES             Location/Qualifiers
     source          1..153147
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8q23"
                     /clone="RP11-79F7"
     misc_feature    1..4
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    153144..153147
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN
```

```
Query Match   58.2%;  Score 667.4;  DB 2;  Length 153147;
Best Local Similarity  99.9%;  Pred. No. 7.7e-173;
Matches  668;  Conservative  0;  Mismatches  0;  Indels  1;  Gaps  0;

Qy     1 ATGGCTCCGGGGCGAAAGCGGAAGCGGGGAGGGGCCCAGGCCAAGAGGCGGCCTCCGGAGATA 60
         ||||||||||||||||| ||||||||||||||||||||||||||||| ||||||||||||||
Db 98486 ATGGCTCCGGGGCGAAAAGCGGAAGCGGGGAGGGGCCCAGGCCAAGAGCGCGCCCTCCGGAGATA 98427

Qy    61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCAGTGGGCGAATGAGAAC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 98426 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCAGTGGGCGAATGAGAAC 98367
```

```
REFERENCE   University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            5 (bases 1 to 18675l)
AUTHORS     Waterston.R.
TITLE       Direct Submission
JOURNAL     Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Nov 8, 2000 this sequence version replaced gi:7630974.
COMMENT     ---------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            --------------- Summary Statistics
                Center project name: H_NH0395G23
            -----------------

            NOTICE: This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Oseogawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-649G15.  Actual start of
            this clone is at base position 1 of RP11-395G23; actual end is at
            base position 18675l of RP11-395G23.

            There are polymorphic base pair differences in the overlap between
            the clone RP11-395G23 and RP11-649G15.
FEATURES             Location/Qualifiers
     source          1..18675l
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-395G23"
                     /clone_lib="RPCI-11"
     repeat_region   90..192
                     /rpt_family="Alu"
     repeat_region   400..514
                     /rpt_family="MIR"
     repeat_region   1104..1401
                     /rpt_family="Alu"
     repeat_region   1402..1577
                     /rpt_family="Alu"
     repeat_region   2954..3028
                     /rpt_family="MIR"
     repeat_region   3283..3501
                     /rpt_family="MER2_type"
     repeat_region   3550..3630
                     /rpt_family="MIR"
     repeat_region   3972..4271
                     /rpt_family="MER1_type"
     misc_feature    4255..4618
                     /note="similar to  EST AA347139  (NID:g1999427)"
     repeat_region   4634..4798
                     /rpt_family="MER1_type"
     repeat_region   5215..5819
                     /rpt_family="MER2_type"
     misc_feature    6093..6323
                     /note="similar to EST AI419746  (NID:g4265677) tg51f06.x1"
     repeat_region   6631..6760
                     /rpt_family="MER1_type"
     repeat_region   6789..6815
                     /rpt_family="MER1_type"
     repeat_region   6816..7300
                     /rpt_family="MER1_type"
     repeat_region   7301..7483
                     /rpt_family="ERV1"
     repeat_region   7488..7657
                     /rpt_family="MER1_type"
     repeat_region   8138..8260
                     /rpt_family="MER1_type"
     repeat_region   8278..8544
                     /rpt_family="L1"
     repeat_region   8640..8787
                     /rpt_family="L1"
     repeat_region   8855..9001
                     /rpt_family="L1"
     repeat_region   9287..9377
                     /rpt_family="MaLR"
     repeat_region   9684..9819
                     /rpt_family="MER1_type"
     repeat_region   10686..10733
                     /rpt_family="L2"
     misc_feature    10725..10947
                     /note="similar to  EST H65488  (NID:g1024228) yr56a02.s1"
     repeat_region   10928..11231
                     /rpt_family="L1"
     repeat_region   11232..11447
                     /rpt_family="Alu"
     repeat_region   11448..11518
                     /rpt_family="Alu"
     repeat_region   11596..11697
                     /rpt_family="L1"
     repeat_region   11899..12354
                     /rpt_family="L1"
     repeat_region   12581..12897
                     /rpt_family="L1"
     repeat_region   13010..13336
                     /rpt_family="Alu"
     repeat_region   13387..14996
                     /rpt_family="MaLR"
     repeat_region   14997..15365
                     /rpt_family="MaLR"
     repeat_region   15470..15708
                     /rpt_family="L1"
     repeat_region   15736..15807
                     /rpt_family="L1"
     misc_feature    15820..17099
                     /note="similar to  EST AW504109  (NID:g7141776)"
     misc_feature    17820..18124
                     /note="similar to  EST AW918394  (NID:g8084170)"
     misc_feature    18314..18470
                     /note="similar to  EST AW318757  (NID:g6748301) un05h10.y1"
     repeat_region   21909..22220
                     /rpt_family="L2"
```

```
repeat_region      /rpt_family="Alu"
                   22662..22964
misc_feature       23594..23647
                   /note="similar to    EST AW318757  (NID:g6748301) un05h10.y1"
misc_feature       23596..23703
                   /note="similar to    EST AW504109  (NID:g7141776)"
misc_feature       23596..23633
                   /note="similar to    EST AW918394  (NID:g8084170)"
repeat_region      24163..24399
                   /rpt_family="L1"
misc_feature       24643..24756
                   /note="similar to    EST AW504109  (NID:g7141776)"
misc_feature       25651..25920
                   /note="similar to    EST AA367453  (NID:g2019771)"
repeat_region      26114..26674
                   /rpt_family="L1"
repeat_region      27351..27417
                   /rpt_family="MIR"
repeat_region      27422..27740
                   /rpt_family="ERVL"
repeat_region      27785..27836
                   /rpt_family="ERVL"
repeat_region      27837..28123
                   /rpt_family="Alu"
```

```
Query Match          58.2%;
Best Local Similarity  99.9%;   Pred. No. 7.7e-173;
Matches   668;   Conservative  0;   Mismatches   0;   Indels   1;   Gaps   0;

Qy       1 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGCCCAGAGAGCGCCCTCCGGAAGATA     60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110228 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGCCCAGAGAGCGCCCTCCGGAAGATA 110169

Qy      61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATAGAAC    120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110168 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATAGAAC 110109

Qy     121 AGGATCAGGCAGGCCCAGGAGGCTGGCTGCCGGGAGGGGACCCAGGACTCACCT    180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110108 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGCCGGGAGGGGACCCAGGACTCACCT 110049

Qy     181 CAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCCAAAG    240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  110048 CAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCCAAAG 109989

Qy     241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAG    300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109988 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAG 109929

Qy     301 GTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA    360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109928 GTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 109869

Qy     361 GGAGGGGACGTGAGCCACCTCAGCCACCAGAGTACAGAGAGGGATGCTGGTGTGCTTGAACCT    420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109868 GGAGGGGACGTGAGCCACCTCAGCCACCAGAGTACAGAGAGGGATGCTGGTGTGCTTGAACCT 109809

Qy     421 GGGCAGCCAGAGATGACATTGACAGATTCCTCCACAGCCACGCCACGCGGAGG    480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109808 GGGCAGCCAGAGATGACATTGACAGAATCCTCCACAGCCACGCCACGCGGAGG 109749

Qy     481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGGAGTGATGGAGCAGGAG    540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109748 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGGAGTGATGGAGCAGGAG 109689

Qy     541 GAGCCCACATGGAGGAGTGACAGCAGGTAGACACAGAGGACAGCCGGCTATGGAGGAGAGCT    600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109688 GAGCCCACATGGAGGAGTGACAGCAGGTAGACACAGAGGACAGCCGGCTATGGAGGAGAGCT 109629

Qy     601 GAGGAGGCCGAGCAGGAGTGACAGGATGAGTGCAGGTGCAGGTGGTGGTCAGGATCAAGCGCCCCTTG    660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  109628 GAGGAGGCCGAGCAGGAGTGACAGGATGAGTGCAGGTGCAGGTGGTGGTCAGGATCAAGCGCCCCTTG 109569

Qy     661 CCCTCCCAG    669
           |||||||||
Db  109568 CCCTCCCAG 109560
```

```
RESULT 11
AC138715
LOCUS       AC138715            164806 bp    DNA     linear   ROD 25-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-283E23 from chromosome 15, complete
            sequence.
ACCESSION   AC138715
VERSION     AC138715.3  GI:38093886
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 164806)
  AUTHORS   Bauer,H., Shah,N., Kozlowicz,A., Haglund,K. and Haakenson,W.
  TITLE     The sequence of Mus musculus BAC clone RP23-283E23
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 164806)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 164806)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 164806)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-FEB-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 164806)
  AUTHORS   Wilson,R.K.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-OCT-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 164806)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Oct 31, 2003 this sequence version replaced gi:2776802.
COMMENT     --------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            --------------- Summary Statistics
            Center project name: M_BA0283E23

            NOTICE:   This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
```

http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (http://www.chori.org) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is overlapped by AC129212.

```
FEATURES        Location/Qualifiers
     source          1..164806
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP23-283B23"
                     /clone_lib="RPCI-23"
     repeat_region   1101..1446
                     /rpt_family="ERVK"
     repeat_region   2103..2280
                     /rpt_family="MaLR"
     repeat_region   2327..2557
                     /rpt_family="MaLR"
     repeat_region   2619..2921
                     /rpt_family="MaLR"
     repeat_region   7357..7639
                     /rpt_family="B4"
     repeat_region   7931..7983
                     /rpt_family="ERV1"
     repeat_region   8120..8275
                     /rpt_family="B4"
     repeat_region   8721..8892
                     /rpt_family="MER1_type"
     misc_feature    10701..10750
                     /note="Sequence derived from PCR product of project DNA."
     repeat_region   12176..12626
                     /rpt_family="L1"
     repeat_region   12619..12762
                     /rpt_family="L1"
     repeat_region   12782..14018
                     /rpt_family="L1"
     repeat_region   14015..14378
                     /rpt_family="L1"
     repeat_region   16341..16746
                     /rpt_family="L1"
     unsure          16732..17658
                     /note="Unresolved simple sequence repeat."
     repeat_region   19633..19901
                     /rpt_family="RMER3"
     repeat_region   20357..20701
                     /rpt_family="B4"
     repeat_region   20715..20891
                     /rpt_family="L1"
     repeat_region   20923..21482
                     /rpt_family="ERVL"
     repeat_region   21626..21857
                     /rpt_family="L1"
     repeat_region   22261..22359
                     /rpt_family="ERVL"
     repeat_region   22401..22524
                     /rpt_family="L1"
     repeat_region   24048..25269
                     /rpt_family="B4"
     repeat_region   25258..25849
                     /rpt_family="L1"
     repeat_region   26538..26711
                     /rpt_family="L1"
     repeat_region   26827..27411
                     /rpt_family="L1"
     repeat_region   27441..27809
                     /rpt_family="L1"
     repeat_region   29246..29368
                     /rpt_family="L1"
     repeat_region   29901..30041
                     /rpt_family="MaLR"
     repeat_region   30666..31622
                     /rpt_family="L1"
     repeat_region   31797..31944
                     /rpt_family="Alu"
     repeat_region   33501..34037
                     /rpt_family="L1"
     repeat_region   34041..34969
                     /rpt_family="L1"
     repeat_region   34966..36100
                     /rpt_family="L1"
     repeat_region   36340..36625
                     /rpt_family="B4"
     repeat_region   37288..37356
                     /rpt_family="MER1_type"
     repeat_region   37631..37806
                     /rpt_family="B2"
     repeat_region   37814..37884
                     /rpt_family="ERV1"
     repeat_region   38610..38975
                     /rpt_family="MaLR"
     repeat_region   39330..39523
                     /rpt_family="B2"
     repeat_region   39531..39713
                     /rpt_family="B2"
     repeat_region   40017..40410
                     /rpt_family="MaLR"
     repeat_region   41037..41178
                     /rpt_family="Alu"
     repeat_region   41654..41775
                     /rpt_family="L1"
     repeat_region   43640..43844
                     /rpt_family="MaLR"
     repeat_region   43820..43974
                     /rpt_family="ERVK"
     repeat_region   43936..44110
                     /rpt_family="MaLR"
     repeat_region   44510..45604
                     /rpt_family="Alu"
     repeat_region   45605..45641
                     /rpt_family="B4"
     repeat_region   46695..46817
                     /rpt_family="MER1_type"
     repeat_region   47790..47898
                     /rpt_family="Alu"
     repeat_region   47901..48444
                     /rpt_family="L1"
     repeat_region   48471..48998
                     /rpt_family="L1"
     repeat_region   49877..49978
                     /rpt_family="L1"
     repeat_region   50038..50318
                     /rpt_family="L1"
     repeat_region   50531..51948
                     /rpt_family="L1"
     repeat_region   51947..52281
                     /rpt_family="L1"
```

```
Query Match      36.6%;   Score 419.8;   DB 10;   Length 164806;
Best Local Similarity   78.3%;   Pred. No. 1.8e-104;
Matches 524;   Conservative   0;   Mismatches 127;   Indels 18;   Gaps  1;

Qy       1 ATGGCTCCGGCGAAAAGGAAAGCGGGGAGGGCCCAGCCAAGACGCCCTCCGGAAGATA 60
           |||||||||| ||||||||||| ||||||||||||||||||||||||||||||||||
Db  161813 ATGGCTCCAGGAGAAAAGGAGAAAAGGGAGGAGGCGGGGGGCCGGCCAAGAGTGCCCTCCGGAAGGTC 161872

Qy      61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAC 120
           ||||||||||||||||||||||||||||||||||||||||||
```

```
REFERENCE   5  (bases 1 to 174846)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (07-MAY-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 174846)
AUTHORS     Wilson,R.
TITLE       Direct Submission
JOURNAL     Submitted (27-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On May 7, 2003 this sequence version replaced gi:28951344.
COMMENT     -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
            --------------- Summary Statistics
            Center project name: M_BB0373G21
            --------------
NOTICE:     This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

SOURCE INFORMATION:
            The RPCI-24 BAC library has been constructed by Pieter de Jong and
            coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
            and/or brain genomic DNA. The clone and detailed information can be
            obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1..174846
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP24-373G21"
                     /clone_lib="RPCI-24"
     repeat_region   2..127
                     /rpt_family="L1"
     repeat_region   557..655
                     /rpt_family="L1"
     repeat_region   673..745
                     /rpt_family="L1"
     repeat_region   878..1194
                     /rpt_family="L1"
     repeat_region   1180..1238
                     /rpt_family="MaLR"
     repeat_region   1255..1315
                     /rpt_family="Alu"
     repeat_region   1318..1718
                     /rpt_family="MaLR"
     repeat_region   2347..2571
                     /rpt_family="MaLR"
     repeat_region   3753..3849
                     /rpt_family="ERVK"
```

```
Db  161873  CGCACAGCAACCCTGGTTATCAATTTGGCCGGAGGTTGGCAGCAGTGGGCGAATGAGAAC  161932
                   |||||||||||||||  |||||||||||||||||||||||||||||||||||
Qy     121  AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGTCGCCGGGAGGGACCCAGGACTCACCT  180
Db  161933  AGTACCAAACAGGCCCAGGAGCCTGCAGCGTGGCTGGCTGCCGGGAGCAACTCATGACCTACCT  161992
                   ||||||||| |||||||||||| | ||||| ||| ||||
Qy     181  CAAGCTCCTAAACCAATCACACCCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCCAAAG  240
Db  161993  AACGGTCTCTAAAGAGCCGGTCC--------------TTACCAGCATGCCCCCAAA  162034
                   ||||||||                      |||||||||||||
Qy     241  TCGCCACCCCGCCTGCCAGAAGACAGCATGGAGCATGGAGCATGGAGAAGCTTCCTGAG  300
Db  162035  ACTCTGTCTCCAAAGCCAGATCGGACGGAGCGAGCAGCACACTCAGAAGAAGCCACCGAG  162094
                   |||||| ||||||||   ||   ||| |||| ||||| |||| ||||||||
Qy     301  GTTTCTCACATCAAAAAGAAAAGAGAGTGTCCAAAACGTGGTCAGCGAAGACTTACGAGAGA  360
Db  162095  GTCTCCCACATTAAAAGGAAAAGGAAAGAGAGTGACCAGAACGGTTGTCAGCAAGGCTTTATGAGAGAGG  162154
                   ||||||||||   |||| |||| ||| |||||||| |||| |||||||
Qy     361  GGAGGGGACGTGAGCCACCTCAGCCACCAGGTACGAGAGGGATGCTGGTGTGTCTTGAACCT  420
Db  162155  GGAGAGGATGTGAACTACCTGAGCCACCAGGTATGAGAATGATGGCGGGCGTGTCTGAAGCT  162214
                   |||| ||| ||||| |||||||||||||||||| ||| || |||| | ||||| |||| ||
Qy     421  GGGCAGCCAGAGAATGACATTGACAGAATCTCCACAGCCACCACGCGTCCCCAACGCGGGAG  480
Db  162215  ATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCACGACTCGGCCAACACGGAGA  162274
                   ||||||||||||||||||||||||||||||  ||| ||||||| |||| |||| |||
Qy     481  AGAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGGTGGAGAGTGATGGAGCAGGAG  540
Db  162275  AGAAAATGCACCAACCTGGTGTCTGAGCTGACCAAAGGGTGGAAAGTGATGGAACAGGAA  162334
                   ||||| ||| |||||||||||||||||| ||||| |||||| ||||||||| |||||
Qy     541  GAGCCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACACGGCTATGGAGGAGAGGCT  600
Db  162335  GAGCCCACGTGGAAGAGTGACAGCGTAGACACAGAGGACACAGAGTGGCTACGGAGGGGGATATG  162394
                   |||||| |||||  |||||||||||||||||||||||||||| |||| |||| ||||
Qy     601  GAGGAGAGGCCCGAGCAGCAGGATGGAGTGCAGGTGGCTGTGTTCAGGATCAAGCGCCCCTTG  660
Db  162395  GAGGAGGGCTGAGCAAGATGCAGCGCCGCTGTGGGCTGCTGCAGGATCAAACGCCCCCTTG  162454
                   ||||||| || |||| |||| ||| | ||||| ||||| ||||||||| |||||||||
Qy     661  CCCTCCCAG  669
            ||||| ||
Db  162455  CTCTCCCCAG  162463
```

```
RESULT 12
AC129212
LOCUS       AC129212              174846 bp    DNA     linear   ROD 27-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-373G21 from chromosome 15, complete
            sequence.
ACCESSION   AC129212
VERSION     AC129212.4  GI:30409922
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 174846)
AUTHORS     Wang,C. and Cotton,M.
TITLE       The sequence of Mus musculus BAC clone RP24-373G21
JOURNAL     Unpublished (2001)
REFERENCE   2  (bases 1 to 174846)
AUTHORS     Wilson,R.
TITLE       Sequencing of Mus musculus
JOURNAL     Unpublished (2001)
REFERENCE   3  (bases 1 to 174846)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (27-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 174846)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (14-MAR-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
```

repeat_region	/rpt_family="L2" 3884..4060
repeat_region	/rpt_family="B2" 4749..5650
repeat_region	/rpt_family="L2" 5620..5828
repeat_region	/rpt_family="L2" 5857..5985
repeat_region	/rpt_family="L1" 5998..6339
repeat_region	/rpt_family="L1" 6620..6713
repeat_region	/rpt_family="L1" 6836..7190
repeat_region	/rpt_family="L2" 7528..7663
repeat_region	/rpt_family="L1" 7667..8059
repeat_region	/rpt_family="L1" 8246..8419
repeat_region	/rpt_family="B2" 8252..8322
tRNA	/product="tRNA-Ser" /note="Likely pseudogene (HMM Sc=22.54 / Sec struct Sc=3.63)" 8424..8491
repeat_region	/rpt_family="L1" 9118..9314
repeat_region	/rpt_family="B2" 10547..10677
repeat_region	/rpt_family="L1" 10675..10725
repeat_region	/rpt_family="L1" 10756..10861
repeat_region	/rpt_family="L1" 10923..11070
repeat_region	/rpt_family="Alu" 13134..13399
repeat_region	/rpt_family="B4" 16149..16330
repeat_region	/rpt_family="B4" 16942..17052
repeat_region	/rpt_family="Alu" 20120..20171
repeat_region	/rpt_family="ERV1" 22985..23186
repeat_region	/rpt_family="B2" 23201..23445
repeat_region	/rpt_family="L1" 23445..23937
repeat_region	/rpt_family="L1" 24744..25079
repeat_region	/rpt_family="MaLR" 25290..25434
repeat_region	/rpt_family="Alu" 25830..25976
repeat_region	/rpt_family="Alu" 27124..27373
repeat_region	/rpt_family="B4" 28911..29230
repeat_region	/rpt_family="L1" 29468..29531
repeat_region	/rpt_family="Alu" 30738..30841
repeat_region	/rpt_family="ERVK" 33438..33560
repeat_region	/rpt_family="B4" 34462..34610
repeat_region	/rpt_family="Alu" 35658..35750
repeat_region	/rpt_family="B4" 36230..36409
repeat_region	/rpt_family="RMER10B" 36455..36818
repeat_region	/rpt_family="B4" 36922..37050
repeat_region	/rpt_family="L1" 37739..38053
repeat_region	/rpt_family="Alu" 39674..39757
repeat_region	/rpt_family="B4" 40854..40978
repeat_region	/rpt_family="L1" 41748..41811
repeat_region	/rpt_family="MER2_type" 42282..42382
repeat_region	/rpt_family="L1" 43128..44024
repeat_region	/rpt_family="L1" 44103..44461
repeat_region	/rpt_family="L1" 44460..45510
repeat_region	/rpt_family="L1" 45503..46334
repeat_region	/rpt_family="L1" 47154..47284
repeat_region	/rpt_family="B4" 47963..48048

```
Query Match            36.6%;  Score 419.8;  DB 10;  Length 174846;
Best Local Similarity  78.3%;  Pred. No. 1.8e-104;
Matches  524;  Conservative  0;  Mismatches  127;  Indels  18;  Gaps  1;

Qy      1 ATGGCTCCGGCGCAAAAGGAAAGCGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATA 60
          ||||||| ||  ||||||||||||| ||| |||| || ||||||||| || ||||||||
Db  11958 ATGGCTCCAGGAGAAAGGGACGCGGGGCCCCAGGAGGGCCCAAGAGTGCCTCCGGAAGTC 12017

Qy     61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGGTTGGCGGGGAATGAGAAC 120
          |||||||| ||||||| |||||  |||||||||||||||||||||||||||||||||||||
Db  12018 CGCACAGCAACCCTGGTTATCAATTTGGCCCGAGGTTGGCAGGTTGGCGGGGAATGAGAAC 12077

Qy    121 AGCATCAGGCAGGCCCAGGAGCCTACAGCTCGGCTGCGGGGAGGGAGACCCAGGACTCACCT 180
          || | ||| ||||||||||||||| ||| ||||||| |||||||| | |  | |  |  |||
Db  12078 AGTACCAAACAGGCCCAGGAGCCTGCAGGCTGCGCCGGGAGCCAACTCATGACCTACCT 12137

Qy    181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAGAGTGCCCCAAAG 240
          ||  ||||||||      |||  |         ||||| |  || |||  ||||||||
Db  12138 AACGCTCCTAAAGAAGCCGGTCC----------TTACCAGCATGCCCCCCAAA 12179

Qy    241 TCGCCACCCGCCTGCGCAAGGACATGCGACAAGGTCGAGAAAGCCCCTGAG 300
          | | |  |||||| | ||||||| |||||||  || |||| || |||| ||
Db  12180 ACTCTGTCTCCAAAGCCAGATCGAGACGGAGGGAGCACACTCAGAAGAGGCCACCGAG 12239

Qy    301 GTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAAGACTTACGAGAGA 360
          || |||||||| |||| ||||||||| || |  |||||| ||| | || |  |||| |
Db  12240 GTCTCCCACATTAAAAGGAAAGAGGTGACCAGGACGGTTGTCAGCAAGGCTTATGAGAGG 12299

Qy    361 GGAGGGGACGTGAGCCACCTCAGCCAGGTACGGACGGATGCTGGTTGGTTGCTTGAACCT 420
          ||||| | ||| | | ||||  ||| ||||  | | |||| ||| ||| || | ||| ||
Db  12300 GGAGGAGTGTGAACTACCTGAGCCCACAGGTATGAGAATGATGGCGGCGGTGTCTGAGGCT 12359

Qy    421 GGGCAGCCAGAAGATGACATTGACAGAATCCTCCAGCCACGGCTCCCCAACCGCGGAGG 480
          |  ||||||||| |||||||||||||||||| |    | | | | |  ||||  ||| |
Db  12360 ATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCAGACTCGACTCACCACGCGAGA 12419

Qy    481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGGAGCAGGAG 540
          ||||||| | |||||||||||||||||||  ||||  ||||||  ||| |  | || |
Db  12420 AGAAAATGCACCAACCTGGTGTCTGAGCTGACCAAAGGCTGGAAAGTGATGAGAACAGGAA 12479

Qy    541 GAGCCCACCATGGACGGAGTGACAGCGTAGACACGAGGACACGAGGACAGCCGCTATGGAGAGGAGGCT 600
          |||||||| ||| |  |||||||| | ||||||  ||||||||||||||  |  |  | |||  |  |
Db  12480 GAGCCCACGTGGAAGAGTGACACGCTAGACACAGAGGACACGAGGACAGTGCTACCGGAGGGGGATATG 12539

Qy    601 GAGGAGAGGCCCGAGCAGCAGGATGGGCTGTCGTCAGGATCAAGCGCCCCTTG 660
```

Db   12540 GAGAGAGAGGCTGAGCAAGATGCAGCGGCCTGTGCGTCCTGCCAGATCAAACCGCCCCTG 12599

Qy   661 CCCTCCCAG 669
              |||||||||
Db   12600 CTCTCCCAG 12608

RESULT 13
AC099736/c
LOCUS   AC099736   186137 bp   DNA   linear   ROD 30-SEP-2003
DEFINITION Mus musculus chromosome 15, clone RP23-228F5, complete sequence.
ACCESSION   AC099736
VERSION   AC099736.10 GI:30841061
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 186137)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE   Mus musculus chromosome 15, clone RP23-228F5
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 186137)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Peterson,K., Rieback,M., Riley,N., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (19-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 186137)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Lui.A., Mabbitt,R., MacLean,C.,
            Lindblad-Toh,K., Liu,G., Lui,A., Landers,T., Levine,R.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (01-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4 (bases 1 to 186137)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui.X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (17-MAY-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5 (bases 1 to 186137)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui.X., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-SEP-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 17, 2003 this sequence version replaced gi:30270708.
            All repeats were identified using RepeatMasker:
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            Smit, A.F.A. & Green, P. (1996-1997)
            --------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------------- Project Information
            Center project name: Li7149
            Center clone name: 228_F_5

FEATURES   Location/Qualifiers
     source   1..186137
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="15"
                     /map="15"
                     /clone="RP23-228F5"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1..6
                     /note="clone_boundary
                     clone end:SP6 end
                     site:EcoRI"

```
Db   127967 ||||||||| ||||||||||  |||||||||||||||||||||||||||||||||||||||||||
            CGCACAGCAACCCTGGTTATCAATTGGCCCGAGGGTTGGCAGCGAGGTGGCGCGGCGAATGAGAAC   127908

Qy      121 AGCATCAGGCAGCCCAGGAGCCTACAGGTGGCTGCGCCGGAGGAGGACCCAGGACTCACCT   180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127907 AGTACCAAAACAGGCCCAGGAGCCTGCAGGTGGCTGCGCCGGGAGCAACTCATGACCTACCT   127848

Qy      181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAG   240
            ||| |||||||||||
Db   127847 AACGCTCCTAAAGAGAGCCGGTCC----------------TTACCAGCATGCCCCCAAA   127806

Qy      241 TCGCCACCCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAGCCCTGAG   300
            ||||
Db   127805 ACTCTGTCTCCAAAGCCAGATCGAGACGGAGGACAACACTCAGAAGAAGCCACCGAG   127746

Qy      301 GTTTCTCACACATCAAAAGAAAGAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA   360
            |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   127745 GTCTCCACCATTAAAAGAAAGAGAGGTGACCCAGCAACCGGTTGTCAGCAAGGCTTATGAGAGG   127686

Qy      361 GGAGGGGACGTGAGCCACCTCAGCCACCAGGATGCGGATGGACGTGGTGTTGCTTGAACCT   420
            ||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||
Db   127685 GGAGGAGATGTGAACTACCTGAGCCCACCAGGGATGAAATGATGGGCGCGTGTCTGAAGCT   127626

Qy      421 GGGCAGCCAGAAGAATGACATTGACAGAATCCTCCACAGCCACGCCCCCACCGCCGGAGG   480
            ||||||||||||||||||||||||||| ||||||||||||||||||| ||||
Db   127625 ATTCAGCCAGAAGAATGACATTGACAGAATCCTTCTTAGTCACGACTCGCCAACACGGAGA   127566

Qy      481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGACTGATCGAGCAGGAG   540
            |||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   127565 AGAAAATGCCACCAACCTGGTGTCTGAGCTGACCAAGGGCTGGAAAGTGATGATGAACAGGAA   127506

Qy      541 GAGCCCACATGGAGAGTGACAGCGTAGACACAGAGGACACAGCGGCTATGGAGAGAGGCT   600
            ||||
Db   127505 GAGCCCACGTGGAGAGAGTGACAGCGTAGACACAGAGGACACAGTGGCTACGGAGGGGATATG   127446

Qy      601 GAGGAGGAGGCCCGAGCAGGAGATGGAGTGCAGGTGCGTGTGGATCAAGCGCCCCTTG   660
            ||||
Db   127445 GAGCCCACGTGGAGAGAGTGACAGCGTAGACACAGAGGACACAGTGGCTCCTGCCAGGATCAAACGCCCCTTG   127386

Qy      661 CCCTCCCAG   669
            |||||||||
Db   127385 CTCTCCCAG   127377
```

RESULT 14
AC136853/c

```
          Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
          Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
          Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
          Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
          Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
          Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
          Karpathy,S., Kelly,S., Kelly,K., Khan,Z., King,L., Kovar,C.,
          Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
          Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
          Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
          Maheshwari,M., Mahindadtne,M., Mahmoud,M., Malloy,K., Mangum,A.,
          Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
          Mawhiney,S., Mcleod,M., Mcneill,T., Meenen,E., Milosavljevic,A.,
          Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K.,
          Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D.,
          Newron,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G.,
          Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H.,
          Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A.,
          Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E.,
          Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y.,
          Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A.,
          Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S.,
          Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A.,
          Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E.,
          Song,X.-Z., Sorelle,R., Sosa,J., Steinle,M., Strong,R., Sutton,A.,
          Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S.,
          Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D.,
          Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J.,
          Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R.,
          Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S.,
          Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X.,
          Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R.,
          Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 73341)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-NOV-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA

COMMENT
          ------------------ Genome Center
             Center: Baylor College of Medicine
             Center code: BCM
             Web site: http://www.hgsc.bcm.tmc.edu/
             Contact: hgsc-help@bcm.tmc.edu
          ------------------ Project Information
             Center project name: KCVW
             Center clone name: CH230-263K14
          ------------------ Summary Statistics
             Sequencing vector: Plasmid;
             Chemistry: Dye-terminator Big Dye: 100% of reads
             Assembly program: Phrap; version 0.990329
             Consensus quality: 28037 bases at least Q40
             Consensus quality: 29795 bases at least Q30
             Consensus quality: 31083 bases at least Q20
          ------------------
     * NOTE: Estimated insert size may differ from sequence length
     *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
     * NOTE: This is a 'working draft' sequence. It currently
     *       consists of 40 contigs. The true order of the pieces
     *       is not known and their order in this sequence record is
     *       arbitrary. Gaps between the contigs are represented as
     *       runs of N, but the exact sizes of the gaps are unknown.
     *       This record will be updated with the finished sequence
     *       as soon as it is available and the accession number will
     *       be preserved.
     *        1       1192: contig of 1192 bp in length
     *     1193       1292: gap of unknown length
     *     1293       2507: contig of 1215 bp in length
     *     2508       2607: gap of unknown length
     *     2608       4113: contig of 1506 bp in length
     *     4114       4213: gap of unknown length
     *     4214       5289: contig of 1076 bp in length
     *     5290       5389: gap of unknown length
     *     5390       6526: contig of 1137 bp in length
     *     6527       6626: gap of unknown length
     *     6627       7760: contig of 1134 bp in length
     *     7761       7860: gap of unknown length
     *     7861       9123: contig of 1263 bp in length
     *     9124       9223: gap of unknown length
     *     9224      10318: contig of 1095 bp in length
     *    10319      10418: gap of unknown length
     *    10419      11713: contig of 1295 bp in length
     *    11714      11813: gap of unknown length
     *    11814      13025: contig of 1212 bp in length
     *    13026      13125: gap of unknown length
     *    13126      14322: contig of 1197 bp in length
     *    14323      14422: gap of unknown length
     *    14423      15639: contig of 1217 bp in length
     *    15640      15739: gap of unknown length
     *    15740      16850: contig of 1111 bp in length
     *    16851      16950: gap of unknown length
     *    16951      18316: contig of 1366 bp in length
     *    18317      18416: gap of unknown length
     *    18417      20103: contig of 1687 bp in length
     *    20104      20203: gap of unknown length
     *    20204      21733: contig of 1530 bp in length
     *    21734      21833: gap of unknown length
     *    21834      23180: contig of 1347 bp in length
     *    23181      23280: gap of unknown length
     *    23281      25124: contig of 1844 bp in length
     *    25125      25224: gap of unknown length
     *    25225      26789: contig of 1565 bp in length
     *    26790      26889: gap of unknown length
     *    26890      28009: contig of 1120 bp in length
     *    28010      28109: gap of unknown length
     *    28110      29909: contig of 1800 bp in length
     *    29910      30009: gap of unknown length
     *    30010      31356: contig of 1347 bp in length
     *    31357      31456: gap of unknown length
     *    31457      33133: contig of 1677 bp in length
     *    33134      33233: gap of unknown length
     *    33234      34430: contig of 1197 bp in length
     *    34431      34530: gap of unknown length
     *    34531      36398: contig of 1868 bp in length
     *    36399      36498: gap of unknown length
     *    36499      37783: contig of 1285 bp in length
     *    37784      37883: gap of unknown length
     *    37884      40081: contig of 2198 bp in length
     *    40082      40181: gap of unknown length
     *    40182      41870: contig of 1689 bp in length
     *    41871      41970: gap of unknown length
     *    41971      44027: contig of 2057 bp in length
     *    44028      44127: gap of unknown length
     *    44128      46329: contig of 2202 bp in length
     *    46330      46429: gap of unknown length
     *    46430      47928: contig of 1499 bp in length
     *    47929      48028: gap of unknown length
     *    48029      50367: contig of 2339 bp in length
     *    50368      50467: gap of unknown length
     *    50468      52413: contig of 1946 bp in length
     *    52414      52513: gap of unknown length
     *    52514      55446: contig of 2933 bp in length
     *    55447      55546: gap of unknown length
     *    55547      57927: contig of 2381 bp in length
     *    57928      58027: gap of unknown length
     *    58028      59802: contig of 1775 bp in length
     *    59803      59902: gap of unknown length
     *    59903      62792: contig of 2890 bp in length
     *    62793      62892: gap of unknown length
     *    62893      66868: contig of 3976 bp in length
     *    66869      66968: gap of unknown length
     *    66969      70615: contig of 3647 bp in length
     *    70616      70715: gap of unknown length
     *    70716      73341: contig of 2626 bp in length.
```

```
FEATURES          Location/Qualifiers
     source       1..73341
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-263K14"
ORIGIN

Query Match      32.5%;     Score 373; DB 2;   Length 73341;
Best Local Similarity   85.6%;     Pred. No. 1.5e-91;
Matches    415;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     662 CCTCCCAGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCAACAGAAATATAGCC 721
               ||||||||| ||||||||||||| ||||||||||||||||||||||||||||||||||
Db   64707 CTTCTAGGGCAAACAGATACTCAGAGACACTCAACTGTAAGGCCCATCGGAAATACAGCC 64648

Qy     722 CAGTGGGCAACTTGAAAGGAGGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCAGA 781
               ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   64647 AAGTGGACAACTTGAAAGGCAGGCAGCAGTGGGCCGATGAACACATACAGTCACAGA 64588

Qy     782 AGCTCAATCCTTTCAGTGAAGAGTTGATTACGGAGCTGGCCATGTCCACCCGCCTACACA 841
               |||||||||||||||||||| |||||||||||||||||| |||||||||||||||||||
Db   64587 AGCTCAACCCCTTCCAGTGATGAATTTGACTATGACCTAGCCATGTCCACTCGACTCCACA 64528

Qy     842 AAGGAGATGAGGGCTATGGCGCCCCAAAGAAGGAAGAAACCAAAACTGCTGAAAGGGCCAAGC 901
               |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   64527 AGGGAGACGAGGGATTATGGCGCCCCCAAAGAGGGAAGCAAGACAGCTGAAAGGGGCCAAGA 64468

Qy     902 GTGCTGAGGAGCACATCTACAGGGAAAATGATGGACATGTGCTTCATTATCTGCACAATGG 961
               ||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||
Db   64467 GAGCCGAGGAGCACATCTATCGGGAAAATATGGGAATTGGCTTTGCTTTGTTATCCGCACAATGG 64408

Qy     962 CTCGCCACAGACGAGGATGGCAAGATCCAGGTTACTTTTGAGATCTCTTTGACAGATACG 1021
               |||||||||||||||||||||||||||||||||||||| ||||||||||||||||| |||
Db   64407 CTCGCCATAGACGAGGATGGCAAGATCCAGGTTACTTTCCGGAGAACTCTTTGATAGADATATG 64348

Qy    1022 TTCGTATTTCAGATAAAGTAGTAGTGGGCCATTCTCATGCGTGCCAGGAAAACATGGACTGGTAG 1081
               |||| |||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   64347 TTCGAATTTCAGATAAAGTGGTGGGCCATCCTCATGCGTGCCAGGAAAACACGGACTGGTGC 64288

Qy    1082 ACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACCGTACTCTCA 1141
               |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   64287 ACTTTGAAGGGGAGATGCTATGGCAAGGCAAGGCAAAAGATGACCATGTTGTGATTACTCTCCTTG 64228

Qy    1142 AGTGA 1146
           |||||
Db   64227 AGTAA 64223

RESULT 15
AC114181
LOCUS       AC114181      220702 bp    DNA    linear   HTG 13-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-19304, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC114181
VERSION     AC114181.5  GI:30579540
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 220702)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenae,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Derano,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,C.M., Finley,M., Flagg,N., Forbes,L., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,J., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hernandez,J.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pal,S., Pasternak,S.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,J., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220702)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220702)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24819086.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----------------- Genome Center
                 Center: Baylor College of Medicine
                 Center code: BCM
                 Web site: http://www.hgsc.bcm.tmc.edu/
                 Contact: hgsc-help@bcm.tmc.edu
            ----------------- Project Information
                 Center project name: GLPO
```

```
        Center clone name: CH230-19304
        ---------- Summary Statistics
        Assembly program: Atlas 3.0;
        Consensus quality: 214139 bases at least Q40
        Consensus quality: 215944 bases at least Q30
        Consensus quality: 217301 bases at least Q20
        Estimated insert size: 225948; sum-of-contigs estimation
        Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
        ---------
*   NOTE: Estimated insert size may differ from sequence length
*       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank draft data.html).
*   NOTE: This is a 'working draft' sequence. It currently
*   consists of 2 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*
*           1  219530: contig of 219530 bp in length
*      219531  219630: gap of unknown length
*      219631  220702: contig of 1072 bp in length.

FEATURES         Location/Qualifiers
     source      1..220702
                 /organism="Rattus norvegicus"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:10116"
                 /clone="CH230-19304"
     misc_feature    1..1179
                 /note="wgs end extension
                 clone_end:Sp6"
     misc_feature    complement(1697..2319)
                 /note="clone_boundary
                 clone_end:Sp6
                 site:EcoRI
                 end_sequence:BH314401"
     misc_feature    complement(218360..218637)
                 /note="clone_boundary
                 clone_end:T7
                 site:EcoRI
                 end_sequence:BH314400"

ORIGIN

Query Match           32.5%;   Score 373;  DB 2;  Length 220702;
Best Local Similarity 85.6%;   Pred. No. 1.5e-91;
Matches  415;   Conservative    0; Mismatches   70;  Indels     0;  Gaps     0;

Qy      662 CCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACACGAAATATAGCC 721
            ||||| |||| ||||||||||||| ||  |||||||||||||| |||| ||| ||||||||
Db   213689 CTTCTAGGGCAAACAGATACTCAGAGACACTCAACTGTAAGGCCCATCGGAAATACAGCC 213748

Qy      722 CAGTGGGCAACTGAAAGGAGGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAGA 781
            ||||||| ||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   213749 AAGTGAGCAACTTGAAAGGCAGGTGGCAGCAGTGGGCCGATGAACACATACAGTCACAGA 213808

Qy      782 AGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACA 841
            |||||||||||||||||||| |||||| ||||| |||||||||||||||||||| |||||
Db   213809 AGCTCAACCCCCTCAGTGATGAATTTGACTATGACCTAGCCATGTCCACTGCACTTCCACA 213868

Qy      842 AAGGAGGATGAGGGCTATGGCCGCCCCCAAAGAAGAGGAACCAAAACTGCTGAAAGGGCCAAGC 901
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   213869 AGGGAGGACGAGGGGATATGGCCGCCCAAAAAGAGGGAGCAAGACAGCTGAAAGGGCCAAGA 213928

Qy      902 GTGCTGAGGAGCACATCACAGGGAAATGATGACATGTGCTTCATTATCTGCACAATGG 961
            ||||||||||||| |||||||| |||||||  || |||  |||||||||||||||||||
Db   213929 GAGCCGAGGAGCACATCATCGGGAAATTATGGAATTGTGTCTTTGTTATCCGGCACAATGG 213988

Qy      962 CTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGGACAGATCCTTTGACAGATACG 1021
            |||| ||||| |||||||| |||||||||| |||||||| |||||||||||  ||||||||
Db   213989 CTCGCCATAGACGAGATCGGCAAGATCCAGGTTACTTTCCGGAGAACTCTTTGATAGATATG 214048

Qy     1022 TTCGTATTCAGATAAAGTAGTGGGCATTCCATGCGTGCCGCGGAGAACATGGGACTGGTAG 1081
            |||| |||||||||||||||||||||| |||||||||||||||||||||||| |||||||
Db   214049 TTCGAATTCAGATAAAGTGGTGGGCATCCTCATGCGTGCCGCGGAGAAACACGGACTGGTGC 214108

Qy     1082 ACTTTGAAGGAGGAGATGCTATGGCAAGGCCGAGATGAGATGCTTGTGTGATTACGCTACTCA 1141
            ||||||||| ||||||||||||||||||||||||||  |||||||||||||| ||||
Db   214109 ACTTTGAAGGGGGAGATGCTATGGCAAGGCAAAGATGACCATGTTGTGATTACTTCTCCTTG 214168

Qy     1142 AGTGA 1146
            |||||
Db   214169 AGTAA 214173

Search completed: May 5, 2004, 02:13:46
Job time : 4467 secs
```

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: May 4, 2004, 08:49:08 ; Search time 64 Seconds
                                       (without alignments)
                                       11299.496 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 2058
Sequence: 1 atggctccggcgaaaagga........tgattacgctactcaagtga 1146

Scoring table: BLOSUM62
           Xgapop 10.0 , Xgapext 0.5
           Ygapop 10.0 , Ygapext 0.5
           Fgapop 6.0 , Fgapext 7.0
           Delop   6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool p/US1064465 9/runat_04052004_084902_28522/app query.fasta_1.1287
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1064465 9 @CGN_1_1_86 @runat_04052004_084902_28522 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LOGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_25:*
        1: sp_archea:*
        2: sp_bacteria:*
        3: sp_fungi:*
        4: sp_human:*
        5: sp_invertebrate:*
        6: sp_mammal:*
        7: sp_mhc:*
        8: sp_organelle:*
        9: sp_phage:*
       10: sp_plant:*
       11: sp_rodent:*
       12: sp_virus:*
       13: sp_vertebrate:*
       14: sp_unclassified:*
       15: sp_rvirus:*
       16: sp_bacteriap:*
       17: sp_archeap:*

No.	Score	%	Length	#	ID	Description
2	1596	77.6	375	11	Q8BU21	Q8buz1 mus musculu
3	1594	77.5	375	11	Q8K431	Q8k431 mus musculu
4	1575	76.5	375	11	Q8K4K7	Q8k4k7 rattus norv
5	1415	68.8	335	11	Q8BLH3	Q8blh3 mus musculu
6	308.5	15.0	162	5	062234	062234 caenorhabdi
7	279.5	13.6	371	5	046052	046052 drosophila
8	228.5	11.1	289	5	Q22133	Q22133 caenorhabdi
9	192	9.3	153	5	Q9VZW6	Q9vzw6 drosophila
10	133.5	6.5	161	10	081881	081881 arabidopsis
11	133.5	6.5	642	10	Q91D34	Q91d34 crypthecodi
12	133.5	6.5	773	16	Q98R04	Q98r04 mycoplasma
13	132	6.4	956	4	Q9UQ39	Q9uq39 homo sapien
14	132	6.4	1262	4	Q9UQ40	Q9uq40 homo sapien
15	132	6.4	2752	4	Q9UQ35	Q9uq35 homo sapien
16	131.5	6.4	95	10	Q8LBN7	Q8lbn7 arabidopsis
17	131.5	6.4	2296	4	Q9UHA8	Q9uha8 homo sapien
18	130.5	6.3	501	4	Q8N677	Q8n677 homo sapien
19	130.5	6.3	501	4	Q9IP5	Q9ip5 homo sapien
20	128.5	6.2	380	2	Q9F692	Q9f692 streptococc
21	128.5	6.2	585	10	Q7XYT7	Q7xyt7 lophopyrum
22	126.5	6.1	742	10	Q7Y0S6	Q7y0s6 taeniatheru
23	126	6.1	605	5	Q965Q9	Q965q9 caenorhabdi
24	126	6.1	608	10	Q84P15	Q84p15 triticum ur
25	126	6.1	836	10	Q8S3T3	Q8s3t3 aegilops ta
26	125.5	6.1	547	4	Q86VQ1	Q86vq1 homo sapien
27	125.5	6.1	811	10	Q94IL5	Q94il5 secale cere
28	125.5	6.1	818	10	Q93XJ8	Q93xj8 triticum ti
29	124.5	6.0	811	10	Q94IJ7	Q94ij7 triticum ae
30	124	6.0	400	2	Q9F689	Q9f689 streptococc
31	124	6.0	803	11	Q8JZQ9	Q8jzq9 mus musculu
32	124	6.0	970	11	Q8CIJ3	Q8cij3 mus musculu
33	123	6.0	971	10	Q8GV12	Q8gv12 triticum ae
34	123	6.0	1049	11	Q7TQH0	Q7tqh0 mus musculu
35	123	6.0	1267	5	Q9VKH3	Q9vkh3 drosophila
36	122.5	6.0	331	5	Q26469	Q26469 schistocerc
37	121.5	5.9	506	16	Q82D70	Q82d70 streptomyce
38	121.5	5.9	836	10	Q8S3T2	Q8s3t2 aegilops ta
39	121.5	5.9	1083	16	086637	086637 streptomyce
40	121	5.9	368	4	Q8WZ58	Q8wz58 homo sapien
41	121	5.9	488	16	Q87W07	Q87w07 pseudomonas
42	121	5.9	926	5	Q9W3G1	Q9w3g1 drosophila
43	121	5.9	1056	11	Q8BIH0	Q8bih0 mus musculu
44	121	5.9	1775	13	Q90XG6	Q90xg6 brachydanio
45	120.5	5.9	601	5	062185	062185 caenorhabdi

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	2029	98.6	381	4	Q8N0Z2	Q8n0z2 homo sapien

ALIGNMENTS

RESULT 1
Q8N0Z2
ID   Q8N0Z2   PRELIMINARY;   PRT;   381 AA.
AC   Q8N0Z2;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   STARS (Hypothetical protein FLJ35375).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22086201; PubMed=11983702;
RA   Arai A., Spencer J.A., Olson E.N.;
RT   "STARS, a Striated Muscle Activator of Rho Signaling and Serum
RT   Response Factor-dependent Transcription.";
RL   J. Biol. Chem. 277:24453-24459(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y...

```
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA   Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA   Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF503617; AAM27268.1; -.
DR   EMBL; AK092694; BAC03948.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   381 AA;  43117 MW;  AC5B8D49A267C46 CRC64;

Alignment Scores:
Pred. No.:                  4.78e-156        Length:        381
Score:                      2029.00          Matches:       381
Percent Similarity:         100.00%          Conservative:  0
Best Local Similarity:      100.00%          Mismatches:    0
Query Match:                98.59%           Indels:        0
DB:                                          Gaps:          0

US-10-644-659A-1 (1-1146) x Q8N0Z2 (1-381)

Qy     1 ATGGCTCCGGGCGAAAAGCGGGAGGGCCCAGCCAAGAGAGCGCCTCCGGAGATA  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle  20

Qy    61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGCGAATGAGAAC  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn  40

Qy   121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGCTGCGCTGCCGGGAGGGACCCAGGACTCACCT  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyTyrGlnAspSerPro  60

Qy   181 CAAGCTCCTAAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAGAGTGCCCCAAAG  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GlnAlaProLysProIleThrProProThrSerHisGlnLysAlaGlnSerAlaProLys  80

Qy   241 TCGCCACCCCGCCCTGCCAGAAGGACATGAGAGTCAGAGGAAAGCCCCTGAG  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 SerProProArgLeuProGluProGluGlyHisGluGlyGlyAspGlyLysSerGluLysAlaProGlu  100

Qy   301 GTTTCTCCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCCAAGAACTTACGACAGAGA  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ValSerHisIleLysLysLysGluVaalSerLysThrValSerLysThrTyrGluArg  120

Qy   361 GGAGGGGACGTGAGCCACCTCAGCGACCTGGTGTGCTTGAACCT  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GlyGlyAspAlaSerHisLeuSerHisArgTyrGluArgAspAlaGlyValLeuGluPro  140

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCACAGCCACCGCTCCCACGCGGAGG  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 GlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisSerGlySerProThrArgArg  160

Qy   481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGAGGCTGGAGAGTGATGAGCAGGAG  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu  180

Qy   541 GAGCCCACATGGAGGAGTGACAGCCTAGACACAGAGGAGACGGCTATGGAGGAGAGGCT  600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GluProThrTrpArgSerAspSerValAspThrGluGluThrGlyMetGluGluAla  200

Qy   601 GAGGAGAGGCCCGAGCAGCAGGATGAGTGCAGGTGCAGTTGGCTGTGGTCAGGATCAAGCGCCCCCTTG  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GluGluArgProGlnGlnAspGluCysArgCysSerTrpLeuValAlaValAlaArgIleLysArgProLeu  220

Qy   661 CCCTCCCAGGTAAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ProSerGlnValAsnArgPheThrGluLysLeuAsnCysLysAlaGlnGlnLysTyrSer  240

Qy   721 CCAGTGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTCATGAACCATACAATCCCAG  780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ProValGlyAsnLeuLysGlyArgTrpGlnGlnTrpGlyHisGluGlnHisIleGlnSerGln  260
```

```
Qy   781 AAGCTCAATCCTTCAGTGAAGAGAGTTTGATTACAGCTGGCCATGTGCCACCCGCCTACAC  840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 LysLeuAsnProPheSerGluGluValLeuIleThrAlaGlyHisValAlaThrArgLeuHis  280

Qy   841 AAAGGAGATGAGGGCTATGGCCGCCCCAAGAACGAAGAACCAAAAACTGCTGAAAGGGCCAAG  900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 LysGlyAspGluG-yTyrGlyArgProLysGluGlyThrLysThrAlaGluLysAlaLys  300

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAATGATGGAAATGATGGACATGTGCTTCATTATCTGCACAATG  960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ArgAlaGluGluGluHisIleTyrArgGluMetMetGluMetMetAspMetCysPheIleIleCysThrMet  320

Qy   961 GCTCGCCACAGACAGAGGCAAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC  1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 AlaArgHisArgArgGlnAspGlyLysIleGlnValThrPheGlyIleSerLeuPheAspArgArgTyr  340

Qy  1021 GTTCGTATTCAGATAAAGTAGTGGCATTCTCATGCGTGCCAGAAAACATGGACTGGTA  1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341 ValArgIleSerAspLysValValGlnIleLysMetArgAlaArgLysHisGlyLeuVal  360

Qy  1081 GACTTTGAAGAGGAGATGCTATGGCAAGGCCGAGATGGACCATGTGTGATTACCGTTACCTACTC  1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AspPheGluGlyGluMetLeuTrpGlnGlyArgAspAspHisGlyArgAspHisValValIleThrLeuLeu  380

Qy  1141 AAG  1143
         |||
Db   381 Lys  381
```

```
RESULT 2
Q8BUZ1
ID   Q8BUZ1          PRELIMINARY;        PRT;    375 AA.
AC   Q8BUZ1;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Weakly similar to F36F2.1 protein.
GN   C130068O12RIK.
OS   Mus musculus (Mouse)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420.563-573 (2002).
DR   EMBL; AK081696; BAC38297.1; -.
DR   MGD; MGI:2444891; C130068O12Rik.
SQ   SEQUENCE   375 AA;  42831 MW;  D55ABB31F80601F1 CRC64;

Alignment Scores:
Pred. No.:                  6.25e-121        Length:        375
Score:                      1596.00          Matches:       301
Percent Similarity:         86.61%           Conservative:  29
Best Local Similarity:      79.00%           Mismatches:    45
Query Match:                77.55%           Indels:        6
DB:                         11               Gaps:          1

US-10-644-659A-1 (1-1146) x Q8BUZ1 (1-375)

Qy     1 ATGGCTCCGGGCGAAAAACGAAAAACGGGAGGGCCCAGCCAAGAGAGCCAAGAGAGCGCCTCCGGAGATA  60
         ||||||||||||| |||:::||
Db     1 MetAlaProGlyGluArgGluArgGlyAlaGlyProAlaLysSerAlaLeuArgLysVal  20

Qy    61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGCGAATGAGAAC  120
         |||||||||||||||||||||||||||||||||||||||||||||||
Db    21 ArgThrAlaThrLeuValIleValIleGluAsnLeuAlaArgGlyTrpGlnLeuTrpAlaAsnGluAsn  40
```

Q8K431
ID   Q8K431   PRELIMINARY;   PRT;   375 AA.
AC   Q8K431;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Striated muscle activator of Rho-dependent signaling.
GN   C130068O12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22086201; PubMed=11983702;
RA   Arai A., Spencer J.A., Olson E.N.;
RT   "STARS, a Striated Muscle Activator of Rho Signaling and Serum
RT   Response Factor-dependent Transcription.";
RL   J. Biol. Chem. 277:24453-24459(2002).
DR   EMBL; AF504061; AAM28877.1; -.
DR   MGD; MGI:2444891; C130068O12Rik.
SQ   SEQUENCE   375 AA; 42823 MW; 2897D1448F24B51C CRC64;

Alignment Scores:
Pred. No.:   9.08e-121   Length: 375
Score: 1594.00   Matches: 300
Percent Similarity: 86.61%   Conservative: 30
Best Local Similarity: 78.74%   Mismatches: 45
Query Match: 77.45%   Indels: 6
DB: 11   Gaps: 1

US-10-644-659A-1 (1-1146) x Q8K431 (1-375)

Qy   1 ATGGCTCCGGGGAAAAGGAAAGCGGGAGGGGCCAGCCAAGAGCGCCTCCGGAAGATA   60
          | |||||| ||||| |||| ||||:::|||||| || || || |||||| ||||
Db   1 MetAlaProGlyLysArgLysAlaGlyGlyAlaSerGlnGluArgLeuArgLysVal   20

Qy   61 CGGCACAGCCACCCTGGTCATCAGCTGGCCCGGAGGTTGGCAGCAGTGGCGAATGAGAAC 120
          ||| |||| | || |||| || ||||||||||||:|| ||| || |||||||||
Db   21 ArgThrArgThrLeuValIleAsnLeuAlaArgGlyTrpGlnInTrpAlaAsnGluAsn   40

Qy   121 AGCATCAGGCAGGCCCAGGAGCCTACAGGGTCGCCGGAGGGACCCAGGACTCACCT 180
          | :::|||||||||||||||| |||||||||:||||| ||||||||||||| ||
Db   41 SerThrLysGlnAlaGlnGluProAlaGlyTyrProLeuProGlyAlaThrHisAspLeuPro   60

Qy   181 CAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCCAAAG 240
          |||||| |||||||||||| --------------TyrGlnHisAlaProLys   74
Db   61 AsnAlaProLysGluAlaGlyPro

Qy   241 TCGGCACCCCGCCTGCCAGAAGGACATGAGCAATGGCAGCAAAGCTCAGAGAGAAAGCCCCTGAG 300
          ||::: ||||: ||||||||||||||||||||
Db   75 ThrLeuSerProLysProArgAspArgAspGluGlyGlnGlnHisHisSerGluGluAlaThrGlu   94

Qy   301 GTTTCTCACATCAAAAAAGAAAGAGAGGTGTCCAAAACGGTGGTCAAGACTTACGAGAGA 360
          ||| :::|||||||| ||||||||| |||||||||||||| || ||||| ||||||
Db   95 ValSerHisIleLysLysGluArgGluValThrArgThrValValSerArgLysLysArg 114

Qy   361 GGAGGGGACCTGAGCCACCTCAGCCACCAGGTACGTAGTAAGACTGGTGTGTGCTTGAACCT 420
          ||||||||||||||||||||||| || || |||| ||:: ||||||| || :::||
Db   115 GlyGlyAspValAsnTyrLeuSerHisArgTyrGluAlaSerGlyArgIleuAsnAspSerGluAla 134

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAAGCGCGGAGG 480
          ||| ||||||||||||||||||||||| |||||||| |||||||| |||
Db   135 IleGlnProGluAsnAspIleAspArgIleLeuHisSerHisAspSerProThrArgArg 154

Qy   481 AGAAAATGTCCAACCTGGTGTCTGAGCTAACCAAGGGCTAACAGGGCTGGAGGTGAGCAGGAG 540
          |||||||||||||||||||| ||||||| ||| ||:: |||||||| |||||||||||
Db   155 ArgLysCysThrAsnLeuValSerGluLeuThrLysGlyTrpLysTrpLysValMetGluGlnGlu 174

Qy   541 GAGCCCACCATGAGGATGACAGCAGTGACACAGACAGGAGGAGGCT 600
          |||||||::: |||| ||:::||| || || ||||||||||||:::
Db   175 GluProThrTrpLysSerAspSerValAspThrGluAspGluGluGlyGlyMet 194

Qy   1 AGCATCAGGCAGGCCCAGGAGCCTACAGGGTCGCCGGAGGGACCCAGGACTCACCT 180
          | :::|||||||||||||||| |||||||||:||||| ||||||||||||| ||
Db   41 SerThrLysGlnAlaGlnGluProAlaGlyTyrProLeuProGlyAlaThrHisAspLeu   60

Qy   181 CAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCCAAAG 240
          |||||| |||||||||||| --------------TyrGlnHisAlaProLys   74
Db   61 AsnAlaProLysGluAlaGlyPro

Qy   241 TCGGCACCCCGCCTGCCAGAAGGACATGAGCAATGGCAGCAAAGCTCAGAGAGAAAGCCCCTGAG 300
          ||::: ||||: ||||||||||||||||||||
Db   75 ThrLeuSerProLysProArgAspArgAspGluGlyGlnGlnHisHisSerGluGluAlaThrGlu   94

Qy   301 GGAGGGGACCTGAGCCACCTCAGCCACCAGGTACGTAGTAAGACTGGTGTGTGCTTGAACCT 360
          ||||||||||||||||||||||| || || |||| ||:: ||||||| || :::||
Db   115 ValSerHisIleLysLysGluArgGluValThrArgThrValValSerArgLysLysArg 114

Qy   361 GGAGGGGACCTGAGCCACCTCAGCCACCAGGTACGTAGTAAGACTGGTGTGTGCTTGAACCT 420
          ||||||||||||||||||||||| || || |||| ||:: ||||||| || :::||
Db   115 GlyGlyAspValAsnTyrLeuSerHisArgTyrGluAlaSerGlyArgIleuAsnAspSerGluAla 134

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAAGCGCGGAGG 480
          ||| ||||||||||||||||||||||| |||||||| |||||||| |||
Db   135 IleGlnProGluAsnAspIleAspArgIleLeuHisSerHisAspSerProThrArgArg 154

Qy   481 AGAAAATGTCCAACCTGGTGTCTGAGCTAACCAAGGGCTAACAGGGCTGGAGGTGAGCAGGAG 540
          |||||||||||||||||||| ||||||| ||| ||:: |||||||| |||||||||||
Db   155 ArgLysCysThrAsnLeuValSerGluLeuThrLysGlyTrpLysTrpLysValMetGluGlnGlu 174

Qy   541 GAGGAGAGGCCCGAGCAGGATGCAGGATGGTCAGGTCGGTGCTGTGGTCAGGATCAAGCCGCCCTG 660
          ||||||| |||||||||||||||||||||||||||||||
Db   195 GluGluArgProGluGlnAspAlaAlaProValAlaProAlaArgIleLysArgProLeu 214

Qy   661 CCCTCCAGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC 720
          |||||| |||||:::|||| :::|||||| :::||||||||||||:||||| :::
Db   215 LeuSerGlnAlaAsnArgTyrSerGluThrGluLeuGlnAsnCysLysLysAlaHisArgLysTyrSer 234

Qy   721 CCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGCTGATGAACACATACAATCCCAG 780
          ||||||||| ||||||||||||||| |||| ||| | |||||||||||| |||||
Db   235 GlnValAspAsnLeuLysGlyArgTrpGlnGlnTrpLeuIleAspGluHisHisValGlnInSerGln 254

Qy   781 AAGCTTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCGGCTACAC 840
          ||||||| | ||| ||| ||:: |||||| |||||||| ||:::||||||||||||
Db   255 LysLeuAsnProPheSerGluAspGluPheArgAspTyrArgLeuLeuAlaMetSerThrArgLeuHis 274

Qy   841 AAAGGAGGATGAGGGCTATGGCGGCTATGGGCCAAAGAAGAAAGAAACTGCTAAAGGCCAG 900
          ||||||||||||||| |||||||||||:||||| ||||||||:: ||||| |||||||||
Db   275 LysGlyAspGluGlyTyrGlyGlyTyrGlyLysProLysLysLysThrAlaLysGlyAlaLys 294

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAATGAAAGTGCTTCATTACGACTGCGACAATG 960
          ||||||||||||||||||| |||||| ||||||||| ::: ||||||||| ||||
Db   295 ArgAlaGluGluHisIleTyrArgGluMetLysCysPheValIleAspCysAspAsnGluArgThrMet 314

Qy   961 GCTCGGCCACAGACGAGATGGCAAGATGGCAAGATCCCAGGTTACTTTTGGAGATCTCTTTGACAGATAC 1020
          ||||||||||||||||||| |||||||||||::: ||||| | | ||||||||||||| | |:::
Db   315 AlaArgHisArgArgAspArgGlyLysLysGlyAspValAlaAsnTyrLeuSerHisArgTyrGluAlaAspArgTyr 334

Qy   1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAAACATGGACTGGTA 1080
          ||| || |||| |||||| | | || ||||||| ||||| ||||||||||||||| |
Db   335 ValArgIleSerAspLysValValGlyIleValIleLeuMetArgAlaArgLysHisGlyLeuVal 354

Qy   1081 GACTTTGAAGGAGGATGCTATGGCAAGGACGATGACCATGTTGTGATTACGCTACTC 1140
          ||| ||||||||||||:|| |||||||||||||||||||||| |||||||||||| ||
Db   355 HisPheGluGlyGlyMetLeuTrpGlnGlyTyrGlyAspAspAspHisValValIleuThrLeuLeu 374

Qy   1141 AAG 1143
          :::
Db   375 Glu 375

RESULT 3

```
Qy    1  ATGGCTCCGGGCGAAAAGGAAAGCGGGGAGGGCCCTCCCGAAGATA  60
         ||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MetAlaProGlyGluThrValArgGluAlaGlyProAlaLeuGlnLysVal  20

Qy   61  CGCACAGCCACCCTGGTCATCAGCTTGGCCCAGAGGTGGCAGAGTGGCGAATGAGAAC  120
         ||||||||||||||||||||||||||||||||||| ::::||||||||||||||
Db   21  ArgArgAlaThrLeuValIleAsnLeuAlaArgGlyTrpGlnTrpAlaAsnGluAsn  40

Qy  121  AGCATCAGGCAGGCCCACAGGCCTGATGAACACATCAATCCCAG  180
         ||||||||||||||||||||||||||||||||:::||||||||
Db   41  SerThrArgGlnAlaGlnGluProAlaAspGlyTrpLeuProGlyAlaThrGlnSerGln  60

Qy  181  CAAGTCTCAAACCAATCACACCCCTACTTCACCACAGAAAGCTCAGAGTGCCCAAAG  240
         ||||||||||||||||||||||||||||||||
Db   61  HisThrProLysGlyProGlyPro----------ArgGlnHisAlaProLys  74

Qy  241  TCGCCACCCGCCTGCCAGAAGGACATGGAAGTGCCCAAAGCTGCTGAAAAGCCCTGAG  300
         |||||||||||||||||||||||| ::::||||::::||||||||
Db   75  ProProSerProLysProAspGlyAspArgGluGlyLysArgGlyLysGluGluAlaThrGlu  94

Qy  301  GTTCTCATCAAAAGAAGAGGTGTCCAAAACGTGGTCAGGACTTACGAGAGA  360
         ||||||||||||||||||||||||:::||||||||||||||
Db   95  ValSerHisIleLysArgLysGlyValThrArgArgThrValValSerLysLysTyrGluArg  114

Qy  361  GGAGGGGAGTGAGCCACCTCAGCCACAGGTACGAGGGATGCTGTGTGCTTGAACCT  420
         ||||||||||||||||||||||||||||||:::||||||
Db  115  GlyGlyGluAspValAsnTyrLeuSerHisArgTyrGluHisAspGlyValGlyValSerGluAla  134

Qy  421  GGGAGCCAGAGAATGACATTGACAGAATCCTCCAGCCACCGGCTCCCCAAGCGCGAGG  480
         ||||||||||||||||||||||||||||||:::||||||
Db  135  ValGlnProAspAsnAspIleAspArgIleLeuLeuSerHisAspSerProThrArgArg  154

Qy  481  AGAAAAATGTGCCAACCTGGTGTCGAGCTAACAAGGCTGGAGGAGTGATGGAGCAGGAG  540
         ||||||||||||||||||||||||||||||:::||||||
Db  155  ArgLysCysThrAsnLeuValSerLysLeuThrLysGlyLeuGluGluMetGluGlnGlu  174

Qy  541  GAGCCCACATGGAGGAGTGACAGCGTAGAACAGAGACACAGAGGCCTATGGAGGAGGGCT  600
         ||||||||||||||||||||||||||:::||||||
Db  175  GluProLysTrpLysSerAspSerIleAspThrGluGlnArgThrGlnArgGlyLeuTrpGlyGlyGlyAspMet  194

Qy  601  GAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATCAAGCGCCCCTTG  660
         ||||||||||||||||||||||||||:::||||||
Db  195  GluGluLeuArgProGluGlnIleAspProLeuGlnIleAsnProValAlaGlnGlnValAlaAlaLeuAlaArgGlyIleLysArgProLeu  214

Qy  661  CCCTCCCAGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATAGC  720
         ||||||||||||||||||||||::::||||| :::||||||
Db  215  HisSerGlnAlaAsnArgTyrSerGluThrLeuAsnCysLysAlaHisArgGlyTyrSer  234

Qy  721  CCAGTGGGCAACTTGAAAGGAGATGGCAGCAGTGGGCTGATGAACACATCAATCCCAG  780
         |||||||||||||||||||||||||::::||||||
Db  235  GlnValAspAsnLeuLysGlyArgTrpGlnGlnTrpAlaAspGluHisIleGlnSerGln  254

Qy  781  AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCGCCTACAC  840
         |||||||||||||||||||||||||:::||||||
Db  255  LysLeuAsnProPheSerAspGluPheAspTyrAspLeuAlaMetSerThrArgLeuHis  274

Qy  841  AAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  900
         ||||||||||||||||||||||::::||||||::::||||||
Db  275  LysGlyAspGluGlyGluTyrGlyGlyArgProLysGluSerLysThrAlaGluArgAlaLys  294

Qy  901  CGTGCTGAGGAGCACATCTACAGGGAAATGATGAACATGTGCTTCATTATCTGCACAATG  960
         ||||||||||||||||||||||:::||||||
Db  295  ArgAlaGluGluHisIleTyrArgGluIleMetGluLeuCysPheValIleArgThrMet  314

Qy  961  GCTCCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTGAGATCCTTTGACAGATAC  1020
         |||||||||||||||||||||||::::||||||
Db  315  AlaArgHisArgArgAspGlyLysIleGlnValThrPheGlyValLeuLeuPheAspArgArgTyr  334

Qy 1021  GTTCGTATTTCAGATAAAGTAGTGGCATTCCATGCGTGCCAGGAAACATGGACTGGTA  1080
         ||||||||||||||||||||||:::||||||
Db  335  ValArgIleSerAspLysAspValValGlyValIleLeuMetArgAlaArgLysHisGlyHisGlyLeuVal  354

Qy 1081  GACTTTGAAGGAGGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC  1140
```

```
Qy  601  GAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATCAAGCGCCCCTTG  660
         ||||||||||||||||||||||||||
Db  195  GluGluLeuArgProGluGlnIleAspProLeuGlnIleAsnProValAlaGlnGlnValAlaAlaLeuAlaProLeu  214

Qy  661  CCCTCCCAGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATAGC  720
         |||||||||||||||||||||| ::::||||| :::||||||
Db  215  HisSerGlnAlaAsnArgTyrSerGluProLeuAsnCysLysAlaHisArgGlyTyrSer  234

Qy  721  CCAGTGGGCAACTTGAAAGGAGATGGCAGCAGTGGGCTGATGAACACATCCCAG  780
         ||||||||||||||||||||||||||
Db  235  GlnValAspAsnLeuLysGlyArgTrpGlnGlnTrpAlaAspGluHisIleGlnSerGln  254

Qy  781  AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCGCCTACAC  840
         ||||||||||||||||||||||||||
Db  255  LysLeuAsnProPheSerAspGluPheAspTyrAspLeuAlaMetSerThrArgLeuHis  274

Qy  841  AAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG  900
         ||||||||||||||||||||||||::::||||||
Db  275  LysGlyAspGluGlyGluTyrGlyGlyArgProLysGluSerLysThrAlaGluArgAlaAlaLys  294

Qy  901  CGTGCTGAGGAGCACATCTACAGGGAAATGATGAACATGTGCTTCATTATCTGCACAATG  960
         ||||||||||||||||||||||:::||||||
Db  295  ArgAlaGluGluHisIleTyrArgGluIleMetGluLeuCysPheValIleArgThrMet  314

Qy  961  GCTCCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTGAGATCCTTTGACAGATAC  1020
         ||||||||||||||||||||||::::||||||
Db  315  AlaArgHisArgArgAspGlyLysIleGlnValThrPheGlyValLeuLeuPheAspArgArgTyr  334

Qy 1021  GTTCGTATTTCAGATAAAGTAGTGGCATTCCATGCGTGCCAGGAAACATGGACTGGTA  1080
         ||||||||||||||||||||||:::||||||
Db  335  ValArgIleSerAspLysAspValValGlyValIleLeuMetArgAlaArgLysHisGlyHisGlyLeuVal  354

Qy 1081  GACTTTGAAGGAGGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC  1140
         ||||||||||||||||||||||:::||||||
Db  355  HisPheGluGlyGluMetLeuTrpGlnGlyArgAspThrLysTrpAspAlaAspHisValValIleThrLeuVal  374

Qy 1141  AAG 1143
         :::
Db  375  Glu 375
```

Alignment Scores:
Pred. No.:   3.15e-119   Length: 375
Score: 1575.00   Matches: 296
Percent Similarity: 85.83%   Conservative: 31
Best Local Similarity: 77.69%   Mismatches: 48
Query Match: 76.53%   Indels: 6
DB: 11   Gaps: 1

US-10-644-659A-1 (1-1146) x Q8K4K7 (1-375)

```
Db   355 HisPheGluGlyMetLeuTrpGlnGlyTyrArgAspHisValValIleThrLeuLeu 374
             |||||||||||||||||||||||:::|||||||||||||||||||
Qy  1141 AAG 1143
Db   375 Glu 375

RESULT 5
Q8BLH3
ID   Q8BLH3         PRELIMINARY;    PRT;    335 AA.
AC   Q8BLH3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Weakly similar to F36F2.1 protein (fragment).
GN   C130068O12RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK045186; BAC32252.1; -.
DR   MGD; MGI:2444891; C130068O12Rik.
FT   NON_TER      1      1
SQ   SEQUENCE    335 AA;    38372 MW;    EB15181FAB749212 CRC64;

Alignment Scores:
Pred. No.:                     335
Score:          2.93e-106     1415.00
Percent Similarity:  85.92%    267
Best Local Similarity:  78.30%    Conservative: 26
Query Match:    68.76%           Mismatches:  42
DB:                              Indels:      6
                                 Gaps:        1

US-10-644-659A-1 (1-1146) x Q8BLH3 (1-335)

Qy   121 AGCATCAGGCAGGCCCAGGAGGCCTACAGGCTGGCTGCCGGGAGGGAGCCCAGGACTCACCT 180
              :::||||||||||||||||||||  ||| |||||||
Db     1 SerThrLysGlnAlaGlnGluProAlaGlyTrpLeuProGlyAlaThrHisAspLeuPro 20

Qy   181 CAAGTTCCTAAAACCAATCACCCCCTACTTCACACCAGAAAGTCAGAAGTGCCCCAAAG 240
             |||         ||              -----------TyrGlnHisAlaProLys 34
Db    21 AsnAlaProLysGluAlaGlyPro-------

Qy   241 TCGCCACCCCGGCCTGCCAGAAGGACATGGAAGGATGGACAAAGCTCAGAGAAAGCCCCTGAG 300
              :::||||||        |:||||||| ||||:::||||||||||
Db    35 ThrLeuSerProLysProAspArgArgAsnGlyGluGlyGlnHisSerGluGluAlaThrGlu 54

Qy   301 GTTTCTCACATCAAAAAGAAAGAGAGTGTCCAAAACGTGGTCAGCAAGACTTACGAGAGA 360
              ||||| |||:::||||:::||||||| |||:::||||||
Db    55 ValSerHisIleLeuLysGluValThrArgThrValSerLysAlaTyrGluArg 74

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGGAGGATGCTGGTTGCTTGAACCT 420
              ||:::||||||||||                |||||||||||||
Db    75 GlyGlyAspValAsnTyrLeuSerHisArgTyrGluAsnAspGlyValSerGluAla 94

Qy   421 GGGCAGCCAGAGAATGCACATTGACACAGATCCTCCACAGCCACACGCGTCCCCACGCGGAGG 480
             |||||||       |||||||||||||||:::|||||||||||
Db    95 IleGlnProGluAsnAspIleAspArgIleLeuLeuSerHisAspSerProThrArgArg 114

Qy   481 AGAAAATGTGCCAACCGGTCTCGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG 540
              |||||| |||||||:::|||||||      |||:::||||||||||||
Db   115 ArgLysCysThrArgSerHisLeuValSerGluValLeuThrLys6lyTrpLys6lyTrpLys6lyTyrValMetGluGlnGluGlu 134

Qy   541 GAGCCCACCATGGAAGGATGAGGAGCTGACAGCGTAGACACAGAGGACAGCGGCGTCATGCTAGCGAGGAGGAGGGCT 600
```

```
Db   135 GluProThrTrpLysSerAspSerValAspThrGlyLysAspSerGlyTyrGlyAspMet 154
             |||||||||||||||||:::|||||||||||||||||||||||:::|||||
Qy   601 GAGGAGAGGCCCGAGCAGGATGGAGTGCAGTGGCTGTGGTCAGGATCAAGCGCCCCTTG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   155 GluGluArgProGluGlnAspAlaAlaValProAlaAlaProLeu 174

Qy   661 CCCTCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAAATATAGC 720
             ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db   175 LeuSerGlnAlaSerAsnArgTyrSerGluThrLeuAsnCysLysAlaHisArgLysTyrSer 194

Qy   721 CCAGTGGGCAACTTGAAAGGACATGGCAGCAGTGGGCTGATGACATACAAATCCCAG 780
             ||||||||||||||||||||||||||||||||:::||||||||||
Db   195 GlnValAspAlaAsnLeuLysGlyArgTrpGlnGlnTrpProAlaAspGluHisValGlnSerGln 214

Qy   781 AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 840
             |||||||||||:::|||||||||||||||||||||||:::|||||||||||||
Db   215 LysLeuLeuAsnProPheSerAspGluPheAspTyrAspLeuAlaMetSerThrArgLeuHis 234

Qy   841 AAAGGAGGATGAGGGCTATGGCCGCCCCAAAGAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG 900
             |||||||||||||||||||||||:::||||||||||||||||
Db   235 LysGlyAspGluGlyTyrGlyArgProLysGluGlyLysSerLysThrAlaGluArgAlaLys 254

Qy   901 CGTGCTGAGGAGGACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG 960
             ||||||||||||||||||||||||:::||||||||||||||||||
Db   255 ArgAlaGluGluHisHisIleTyrArgGluIleMetGluLeuCysPheValIleArgThrMet 274

Qy   961 GCTCGCCCAGACGAGATGGCAAGATCCAGTTACTTTGGAGATTCTTTGACAGATAC 1020
             ||||||||||||||||||||||:::|||||||||||||||||
Db   275 AlaArgHisArgArgAspGlyLysHisIleGlnValThrPheGlyGluLeuPheAspArgTyr 294

Qy  1021 GTTCGTATTTCAGATAATAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGTA 1080
             |||||||||:::|||||||||||||||||||||||||||
Db   295 ValArgIleSerAspLysValValGlyIleValLeuMetArgAlaArgLysHisGlyLeuVal 314

Qy  1081 GACTTTGAAGGAGGAGATGCTATGCAAGGCCAGATGACCATGTTGTGATTACGGCTACTC 1140
             ||||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db   315 HisPheGluGlyGluMetLeuTrpGlnGlyTyrArgAspHisValValIleThrLeuLeu 334

Qy  1141 AAG 1143
          :::
Db   335 Glu 335

RESULT 6
O62234
ID   O62234         PRELIMINARY;    PRT;    162 AA.
AC   O62234;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   F36F2.1 protein.
GN   F36F2.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cottage A.;
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81532; CAB04325.1; -.
DR   PIR; T21860; T21860.
DR   WormPep; F36F2.1; CE15975.
SQ   SEQUENCE    162 AA;    18490 MW;    C21116625BBF64B9 CRC64;

Alignment Scores:
```

```
Pred. No.:                    1.42e-16        Length:               162
Score:                        308.50          Matches:               61
Percent Similarity:           70.54%          Conservative:          30
Best Local Similarity:        47.29%          Mismatches:            31
Query Match:                  14.99%          Indels:                 7
DB:                                           Gaps:                   3

US-10-644-659A-1 (1-1146) x 062234 (1-162)

Qy    760 GATGAACAACATACAATCCAG---AAGCTCAATCCTTCAGTGAAGAGTTGATTACGAG 816
                 :::::::||||||||:::|||  ||  ::|||:|||::|||
Db     19 GluGlnAsnValAlaThrGlnSerLysAspAspValTyrSerLysAspPheThrGlnLys 38

Qy    817 CTGGCCATGTCCACCCGCTACACAAAGGAGGATGAGGGCTATGGCCGCCCAAAGAAGGA 876
                 :::::                   |||
Db     39 ---------LysMetAspLysSerSerSerGluTyrGlyArgProLysProGly 53

Qy    877 ACCAAAACTGCTGAAAGGGCCAAGGCGTGCTGAGGAGCACACTACAGGGAAATGATGGAC 936
                 |||  ||||||||:::||||:||||||||||||||:||:||||||
Db     54 ThrLeuThrGluGlnArgAlaLysLysAlaAlaAlaHisValHisArgGluMetLeuThr 73

Qy    937 ATGTGCTTCATTATCTGCACAATGGCTCGCCCACAGACCGAGATGGC---AAGATCCAGGTT 993
                 :::          |||              :::::|||
Db     74 LeuCysGluValValGluAspThrGlyLysGlnIluLysGluGlyAspProIleArgIle 93

Qy    994 ACTTTGGAGATCCTTTGACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCTC 1053
                 ||||||||| ||||| |||| ||||| |::||||
Db     94 ThrPheGlyArgLeuPheThrIleTyrValAsnIleSerAspLysValValGlyThrLeu 113

Qy   1054 ATGCGTGCCAGGAAACATGGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCGA 1113
                 :::|||||     |||||| |||||||||
Db    114 LeuArgAlaArgLysHisLysMetIleAspPheGluGluMetLeuPheLeuGlnLysArg 133

Qy   1114 GATGACCATGTTGTGATTACGCTACTC 1140
                 ||||||||:::||||||
Db    134 AspAspHisValIleIleThrLeuLeu 142
```

RESULT 7
O46052
ID   O46052   PRELIMINARY;   PRT;   371 AA.
AC   O46052;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   EG:152A3.3 protein (HL01173P) (CG3630-PA).
GN   EG:152A3.3 OR CG3630.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=OREGON-R;
RX   MEDLINE=20196011; PubMed=10731137;
RA   Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D., Barrell B.,
RA   Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E., Dreano S.,
RA   Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D., Minana B.,
RA   Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov V.,
RA   Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA   Modolell J., Peter A., Schottler P., Werner M., Mourkioti F.,
RA   Beinert N., Dowe G., Schafer U., Jackle H., Bucheton A.,
RA   Callister D.M., Campbell L.A., Dalamitsou A., Henderson N.S.,
RA   McMillan P.J., Salles C., Tait E.A., Valenti P., Saunder R.D.,
RA   Glover D.M.;
RT   "From sequence to chromosome: the tip of the X chromosome of D.
RT   melanogaster.";
RL   Science 287:2220-2222(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,
RA   Clamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,
RA   Harris N.L., Kronmiller B., Marshall B., Millburn G.H., Richter J.,
RA   Russo S., Searle S.M.J., Smith S.J., Shu S., Smutniak F.,
RA   Whitfield E.J., Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J.,
RA   Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (JAN-2003) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
DR   EMBL; AL009194; CAA15697.1; -.
DR   EMBL; AY060993; AAL28541.1; -.
DR   EMBL; AE003423; AAF45738.3; -.
DR   FlyBase; FBgn0023540; EG:152A3.3.
SQ   SEQUENCE 371 AA; 39430 MW; 3C3B6D2E58FEB36D CRC64;

Alignment Scores:
Pred. No.:   3.68e-14   Length: 371

Score: 279.50 Matches: 64 Mismatches: 49
Percent Similarity: 61.81% Conservative: 25 Indels: 11
Best Local Similarity: 44.44% Mismatches: 50 Gaps: 4
Query Match: 13.58% Indels: 5
DB: Gaps: 2

US-10-644-659A-1 (1-1146) x O46052 (1-371)

Qy 718 AGCCAGTGGCAACTTGAAAGGAGAGGAGGCAGCAGTGGGCTGATGAACACATACAATCC 777
        |||::||| ::: ::. | ||:||| ||||||
Db   17 SerProLeuSerSerLysValAlaMetPheAsnAsnGlnAlaThrGlnHisLysGlnSer 36

Qy 778 CAGAAGTCAATCCTTCAGTGAAGAGTTTGATTACGACGGCCATGTCCACCCGCCTA 837
        :|| |||||||||::||| ::   :::
Db   37 GlnLeuLeuAsnProPheSerGlnAsp---------GlyArgAlaAlaSerProLysPro 53

Qy 838 CACAAAGGAGGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCC 897
         :: ||||||| :||| |||||
Db   54 ThrPheSerLysAspGlnTyrGlyLysProLeuAlaGlySerLeuThrGluMetArgGly 73

Qy 898 AAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTCTTCATTATCCGCACA 957
        :::||| |||::: :|||||||||:||||||||:::|
Db   74 GlnLysAlaAsnIleHisValMetLysGluMetLeuGlnIleLeuCysGlnIleIleAsnSer 93

Qy 958 ATGGGCTCGCCACAGACGAGTGGCAAGATCCAG------GTTACTTTGGAGATCTCTTT 1011
        :::||| ::: ||||:||:|||:
Db   94 GluGlyTyrAspValLysAspGluProThrMetArgValIleProPheGlyGluLeuPhe 113

Qy 1012 GACAGATACGTTCGTATTCGATAAAGTAGTGGGCCATTCTCATGCGTGCCAGGAAAACAT 1071
        ::::||| ||||||||||||||||
Db 114 AsnIleTyrAsnTyrIleSerAspLysValGlyIleLeuLeuLeuArgAlaArgLysHis 133

Qy 1072 GGACTGGTAGACTTTGAAGGAGAGGATGCTATGGCCAAGGCCCGAGATGACCATGTTGTGATT 1131
        :|| ||||||||||||||| ||| :::
Db 134 LysLeuValAspPheGluGlyGluMetLeuTyrGlnArgArgAspAspValProVal 153

Qy 1132 ACGCTACTCAAG 1143
        |||||||||
Db 154 PheLeuLeuLys 157

RESULT 8
Q22133
ID Q22133; PRELIMINARY; PRT; 289 AA.
AC Q22133;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE T04A8.4 protein.
GN T04A8.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Palmer S.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; 235663; CAA84722.1; -.
DR PIR; T24421; T24421.
DR WormPep; T04A8.4; CE01073.
SQ SEQUENCE 289 AA; 32972 MW; 20DB67F731F03098 CRC64;

Alignment Scores:
Pred. No.: 4.83e-10 Length: 289
Score: 228.50 Matches: 56
Percent Similarity: 58.04% Conservative: 27

Best Local Similarity: 39.16% Mismatches: 49
Query Match: 11.10% Indels: 11
DB: Gaps: 5 Gaps: 4

US-10-644-659A-1 (1-1146) x Q22133 (1-289)

Qy 721 CCAGTGGCAACTTGAAAGGG---------AGATGGCAGCAGTGGGCTGATGAACACATA 771
        |||::|||::::   ::::: :::::| :::
Db   2 ProIleGlySerAlaSerAspThrIleArgGlyAsnAlaValAlaGlnAlaAsnGlu 21

Qy 772 CAATCCCAGAAGTCAATCCTTTCAGTGAAGAGTTTGATTACGAGTGCCATGTCCACC 831
        :::||| :|||| |||| ---AlaPheAspThr 40
Db   22 GluValLeuLysLysAsnProTyrSerAspThrTyrLysIleGln--AlaPheAspThr 40

Qy 832 CGCCTACACAAAGGAGGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAA 891
        :::   :::||||||||||:::||||
Db   41 Lys------------AsnTyrGlyArgProProGlySerLysThrGluAla 54

Qy 892 AGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTCTTCATTATC 951
        ||| ::||| ||||||:||||||||||:::||
Db   55 ArgGlyIleLysAlaGlyvalHisValCysArgGluIleLeuPheLeuCysGluThrIle 74

Qy 952 TGCACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTT 1011
        :::||| :::||||:||:||
Db   75 AspSerAsnAlaAspGlyGluGluProHisLys--TyrValLysPheGlyLysLeuPhe 93

Qy 1012 GACAGATACGTTCGTATTCGATAAAGTAGTGGGCCATTCTCATGCGTGCCAGGAAAACAT 1071
        :::   ::::|||:::|||||
Db   94 AsnIleTyrSerPheTyrSerAspLysValGlyMetLeuIleArgAlaArgLysTyr 113

Qy 1072 GGACTGGTAGACTTTGAAGGAGAGGATGCTATGGCCAAGGCCCGAGATGACCATGTTGTGATT 1131
        :|| ||||||||||||||| ||| :::
Db 114 GlyLeuValHisPheGluGlyGluMetLeuTyrGlnArgGlnAspAspGluLysIleIle 133

Qy 1132 ACGCTACTC 1140
        |||||||
Db 134 ThrMetLeu 136

RESULT 9
Q9VZW6
ID Q9VZW6; PRELIMINARY; PRT; 153 AA.
AC Q9VZW6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG2113 protein (AT18037p).
GN CG2113.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

```
Score:                    192.00       Matches:        44
Percent Similarity:       56.20%       Conservative:   24
Best Local Similarity:    36.36%       Mismatches:     49
Query Match:               9.33%       Indels:          4
DB:                                    Gaps:            2

US-10-644-659a-1 (1-1146) x Q9VZW6 (1-153)

Qy   784 CTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAA 843
             ::||||||||||     ::::             |||              :::
Db     3 IleAsnProPheAlaHis-------TyrAsnValAsnGluMetProLysArgThrPhe 19

Qy   844 GGAGATGAGGCTATGGCCCGCCCCAAAGAAGAGAACCAAAACTGCTGAAAGGGCCAAGGGT 903
             ::|||   |||||:::|       :::::    :::  ||||:
Db    20 ProGluGluGlyTyrGlyArgAlaProThrGlySerLeuSerGluArgSerLeuGln 39

Qy   904 GCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATGGCT 963
             |||           :::||||   ||::|| ::||:|||
Db    40 AlaAsnValArgAlaLeuGluGluIleLeuGlnLysAspLeuIleLeuCysSerGly 59

Qy   964 CGCCACAGACGA---GATGGCAAGATCCAGTTACTTTGGAGATCTCTTTGACAGATAC 1020
             ||||||:::        ::  :: ||||||:
Db    60 ArgAspAspProIleAspGlyArgGlyValLeuValLeuAlaPheGlyGlnLeuPheGluThrTyr 79

Qy  1021 GTTCGTATTCAGATAAAGTAGTGGGCATTCCATCGTGCCAGGAAACATGGACTGGTA 1080
             |||:||    ||||||:::::
Db    80 AsnAsnIleSerAspLysLeuLeuAlaThrLeuLeuGlyAlaArgLysTyrGlyPheVal 99

Qy  1081 GACTTTGAAGGAGAGATGCATGGCAAGGGCCGAGATGACCATGTTGTGATTACGGTACTC 1140
             |||   :::   |||||||    ||  |||   |||::
Db   100 AspPheSerGlyGluThrLeuPheGlnGlyArgAspAspAspThrGluProValArgLeuLeu 119

Qy  1141 AAG 1143
         :::
Db   120 Arg 120
```

RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera E., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez C., Houck J., Hoskins R.A., Li P., Mattei B., Moshrefi A.,
RA   Ibegwam C., Jalali M., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   McIntosh T.C., Moy M., Murphy B., Park S., Patel S., Pfeiffer B.,
RA   Pacleb J., Paragas V., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Phouanenavong S., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman J., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield T.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RP   FlyBase;
RA   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003476; AAF47699.2; -.
DR   EMBL; AY113283; AAM29288.1; -.
DR   FlyBase; FBgn0035384; CG2113.
SQ   SEQUENCE 153 AA; 17529 MW; F47F7EE126DE8DF0 CRC64;

Alignment Scores:
Pred. No.:   3.95e-07   Length: 153

RESULT 10
081881
```
ID   081881    PRELIMINARY;    PRT;    161 AA.
AC   081881;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   T16L11.130 OR AT4G33640.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Obermaier B., Deutschenbaur S., Piravandi E., Hoheisel J., Jesse T.,
RA   Heijnen L., Vos P., Mewes H.W., Mayer K.F.X., Schueller C., Bevan M.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Obermaier B., Deutschenbaur S., Piravandi E., Mewes H.W., Lencke K.,
RA   Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RP   EU Arabidopsis sequencing project;
RA   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031394; CAA20577.1; -.
DR   EMBL; AL161583; CAB80081.1; -.
DR   PIR; T04981; T04981.
KW   Hypothetical protein.
SQ   SEQUENCE 161 AA; 18337 MW; 4F42A35D064DE8B5 CRC64;
```

Alignment Scores:
Pred. No.:   0.0221   Length: 161
Score: 133.50   Matches: 35

```
Percent Similarity:     51.06%      Conservative:    13
Best Local Similarity:  37.23%      Mismatches:      33
Query Match:            6.49%       Indels:          13
DB:                     10          Gaps:            2

US-10-644-659A-1 (1-1146) x O81881 (1-161)

Qy   862  CGCCCCAAAGAAGGAACCAAAAACTGCTGAAAGGCCAAGCGTGCTGAGGAGCCACATCTAC  921
          |     ::  :::|||      ||       ||     |||||  |||| ||||| |||
Db    62  LysGluGluGlnArgMetAsnValAspGluIleGlnLysLeuGlnLysGluGlnIleHis   81

Qy   922  AGGGAAAATGATGGACATGTGCTTCATTATCTGCACAATGGCTCGCCAGACCAGATGGC    981
          |||                ::: :        |||        ::: |||||| ||||
Db    82  Arg                          LeuGlySerArgGlnThrAspGly          90

Qy   982  AAGATCCAGGTTACTTTTGGAGATCTCTTT------GACAGATACGTTCGTATTTCAGAT  1035
          ::|||   ||||       ||||        ||  |||||| |||   ||| ||||||
Db    91  SerTyrLysValThrPheGlyValLeuPheAsnAspAspArgCysAlaAsnIlePheGlu  110

Qy  1036  AAAGTAGTGGCATTCTCATGCGTGCCAGGAAACATGGACTTGGTAGACTTTGAAGGAGAG  1095
          |||||||     :|||||   |||||||       ||:|||      ||||||||||||
Db   111  AlaLeuValGlyThrLeuArgAlaAlaLysLysArgLysIleValAlaPheGluGluGlyGlu  130

Qy  1096  ATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGGCTA                 1137
          ::|||    |||||     |||      ||| ||||||||||
Db   131  LeuLeuLeuGlnGlyValHisAspLysValGluIleThrLeu                 144
```

RESULT 11

```
Q9LD34
ID   Q9LD34              PRELIMINARY;    PRT;   642 AA.
AC   Q9LD34;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Dinap1-interacting protein 5.
GN   DIP5.
OS   Crypthecodinium cohnii (Dinoflagellate).
OC   Eukaryota; Alveolata; Dinophyceae; Gonyaulacales; Crypthecodiniaceae;
OC   Crypthecodinium.
OX   NCBI_TaxID=2866;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99305750; PubMed=10377987;
RA   Bhaud Y., Geraud M.L., Ausseil J., Soyer-Gobillard M.O., Moreau H.;
RT   "Cyclic expression of a nuclear protein in a dinoflagellate.";
RL   J. Eukaryot. Microbiol. 46:259-267(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Derelle E., Ausseil J., Moreau H.;
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF25566.; AAF70817.1; -.
DR   InterPro; IPR003409; MORN.
DR   Pfam; PF02493; MORN; 3.
DR   SMART; SM00698; MORN; 3.
SQ   SEQUENCE  642 AA;  66779 MW;  9A6CD34381A5C53F CRC64;
```

```
Alignment Scores:
Pred. No.:              0.028         Length:          642
Score:                  133.50        Matches:         107
Percent Similarity:     38.30%        Conservative:    55
Best Local Similarity:  25.30%        Mismatches:      137
Query Match:            6.49%         Indels:          125
DB:                     10            Gaps:            21

US-10-644-659A-1 (1-1146) x Q9LD34 (1-642)

Qy     4  GCTCCCGGCGAAAAAGGAAAGCGGGAGGGCCCCAGCGGGAAGGAGAGCCCTCCGAAGATACGG    63
          |||||||              :::|||           |||          |:::|||
Db    69  AlaProGlyAlaGly------Lys------LeuProGlyAlaProThrGlyLeuProGl     86

Qy    64  ACAGCCACCCTGGTCATCAGCTTGGCCTGACAGGTTGGCAGTGGCGAATGAGAACAGC    123
          |||       ::: :|||         |||||::::::       :|||||||||||
Db    86  yAlaPro------------LysAlaLeuProLys-----GlyAlaGlySerLeuProThrAl  101
```

```
Qy   124  ATCAGGCAGGCCCAGAGCCTACAGGCTGGCTGCCGGGAGGGAGGGACCCAGGACTCACCTCAA  183
          |||      |||::::::||            |||||          ||||||
Db   101  aProGlyLysAlaProLysAlaLeuProGlyGly--ProGly-                    113

Qy   184  GCTCCTAAACCAATCACACCCCTACTTCACCAGAAAGCTCAGAGTGCCCAAAGTCG     243
          |||||| ::|||                               ||||||
Db   114  AlaGlyArgProThrAlaProPro-----                              ---AlaProLysSer  125

Qy   244  CCACCCCGCCTGCCAGAAGGACATGGAGAT--GGACAAAGCTCAGAGAAAGCCCCTGAG    300
          |||||| :::||          ||||||||    ||     ||||::::||||||||||
Db   126  LeuProSerLeuArgAlaGlyProSerAspThrGlyPheSerAlaProProAlaPro--  144

Qy   301  GTTTCTCACATCAAAAGAAGAGAGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA    360
          :::      ||                          ||
Db   145  ---AsnGlnLeuProAlaAlaProProLysProAlaAlaValProGlyAlaProLysPro  163

Qy   361  GGAGGGGA                     --CGTGAGCCACCTCAGCCACCAGGTACGAGAGGGGATGCTGG  407
          ||||||||                       ::|||:::||
Db   164  GlyGlyLeuProGlyGlyTyrProAlaAlaProLysProGlyValLeuProGly-AlaGl  183

Qy   408  TGTG-                                                       CTTGAACCTGG  422
          ||||                                                        ::|||||||||
Db   183  yValValProProAlaProLysAlaLeuProAlaAlaLysValGluGluValGluProGl  203

Qy   423  GCAGCCAGAGAATGACATTGACAGAATCCTCCAGCGACCACGGCTCCCCAACGCGGAGGAG  482
          ||| ::  :::                        |||
Db   203  uGluProLeuSerGluGluLeuGlyLeuGlyLyGlyGly-                       217

Qy   483  AAAATGTGCCAACCTGGTCTGCTGAGCTAACCAAGGGCTGGAGA--                 525
          |||||                                     :::||
Db   218  -----                                        GlnAsnTrpArgGlnTyrIleGluAlaAs  227

Qy   526  GTGATGGAGCAGGAGGAGGGCCCATGGAGGAGTGACAGCCTAGACAC  572
          |||         :::|||||||||  :::
Db   227  pAlaAspAspGlyIleAspAspAspGluProSerAspValGlyAspGluGlyGlu1  247

Qy   573  AGAGGACAGCGCGGCTGATGGAGGAGGAGGCT-----                    GAGGAGAGGCC  611
          |||||||||||||     :::
Db   247  uGluAspSerGlyGluAspSerAspAlaGluProAlaAlaSerAlaAspGluGlyArg  267

Qy   612  CGAGCAGGATGAGGTGAGTGCAGGTGCCAGGTGGCTGTGGTGCTGGTCAGGATCAAGCGGCCCCTTGCCCTCCCAGT  671
          ::: ::::::|||  ||:::::::||
Db   267  gLysArgAsp-               MetGluLeuLysAlaArgGlnGluThrProGlyGlnTry  284

Qy   672  AAACAGATTTACAGAGAAAC----                                   TCAACTGCAAAGCCCAAC--AGAAATA  716
          ||||||||||||||
Db   284  rAlaArg-AlaGlnMetAsnGluAlaLysSerAlaAlaLysSerProLysProGluThrP  304

Qy   717  TAGCCCAGTGGGCAACT-                                        TGAAAGGAGATGGCAGCA  752
          |||||||||||
Db   304  heAlaAspAlaValThrAspGluGluAlaSerProProHisLysSerAlaGlySerA  324

Qy   753  GTGGGGCTGATGAGCAACACATACAATCCCAGAGCTCAATCCTTTCA-           CA  796
          :::                          ::: :|||
Db   324  spAsnGluValArgGluAspSerAlaAlaPheLeuSerGlnAlaProSerSerL  344

Qy   797  --GTGGAAGAGTTTGATTACGAGCTGGCCATGTCCACCGCCTACACAAAGGAGATGAGG  854
          ::::::::                                :::||
Db   344  ysLysArgSerLeuIleSerArg-                    ValPheSerLysThrLysG  359

Qy   855  CTATCGCCCGCCCAAAGAAGGAACCAAAACTGCTGAAAGGCCAAGCGTGCTGAGGAGCA  914
          |||   ||||||
Db   359  lyGlnAlaSerProSer-                           LeuSerGlyThrSerLeuLeuLys-Ala  373

Qy   915  CATCTACAGGGAAATGAT-                                       GGACATGTGCTTCATTATCTGCACAAT  959
          :::::  :: |||                                            ::|||
Db   374  AsnIleAspSerHisAspGluSerTrpTyrAlaAlaHisLeuSerArgLeuArgGlu  393

Qy   960  GGCTCGCCAGACCAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATA  1019
          |||       ::: :|||
Db   394  GlyLeuProCysThrLys-                           IleGly-ThrAsnGlyLysProTyrAspArgAr  410
```

Qy   1020 CGTTCGTATTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCGAGGAAACATGGA 1074
            ||| ||| ||| :: :: :: ||| || || :: || ::: |||
Db   410 gValHisIleAspAlaArgAsnMetGlnValGluIleArgGlyGlyArgThrGly 428

RESULT 12
Q98R04
ID   Q98R04   PRELIMINARY;   PRT;   773 AA.
AC   Q98R04;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2003 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   LIPOPROTEIN.
GN   MYPU_2060.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445563; CAC13379.1; -.
DR   PIR; F90537; F90537.
DR   MypuList; MYPU_2060; -.
DR   InterPro; IPR000437; Prok_lipoprot_S.
DR   PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW   Complete proteome.
SQ   SEQUENCE   773 AA; 88499 MW; 449343B50BFB1A0 CRC64;

Alignment Scores:
Pred. No.:   0.0289   Length: 773
Score: 133.50   Matches: 80
Percent Similarity: 34.42%   Conservative: 47
Best Local Similarity: 21.68%   Mismatches: 130
Query Match: 6.49%   Indels: 113
DB: 16   Gaps: 15

US-10-644-659A-1 (1-1146) x Q98R04 (1-773)

Qy   112 AATGAGAACAGCATCAGGCAGCGGCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGACCCAG 171
            ||| ||||||| |||
Db   60 AsnArgGlnAsnSerIleSerProGlnAsnPro----------------------------- 70

Qy   172 GACTCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCACCACAGAAAGTCAGAGT 231
            |||||| ::|||||:   :::
Db   71 AspSerSerLysThrProGluThrProGluIleValProProThrLysProGluAsnLys 90

Qy   232 GCCCCAAAGTGGCCACCCCGCCTCCAGAAGGACATGGAAGGATGGAGAAGAAGCTCAGAG-- 288
            |||:::|| ||||||| ||||||:::
Db   91 GluProGlnSer---ProLysAspProLysAlaProGluIleLysAspAsnGlyGlnLysAsnGlyAsnGluGly 109

Qy   289 --AAAGCCCCTGAGGTTTCTACACTACAAAAAGAAAGAGGTGTC------------------ 329
            |||||||:::   :: ||:|||:
Db   110 SerLysAlaProGluIleLysAspMetSerGlnLysAspGlnAlaProGlnAlaProGlnValProGln 129

Qy   329 --- 329

Db   130 LysGlnProGluAspProLysLysProGluThrGlnLysLysProProValLysSerGluAsp 149

Qy   330 CAAAACGGTGGTCAGCAAGACTTACGA------GAGAGGAGGGACGTGAGCCACCTCAG 383
            ||||||| ::: |||:::::
Db   150 GlnAsnLysGluProGlnAspProLysAlaProGluLysGlnAlaGluThrProLysAsp 169

Qy   384 CCACAGGTACGAGAGGGGATGCTCGTGTGCTTGACCTCGGCAGCCAGA----------- 431
            :::
Db   170 ProGlnValLysAsp-------------------------MetAlaAsnLysAsnIleGlnGly 182

Qy   432 -----------GAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAG 479

Db   183 ProLysValProGluLysGlnAlaGlnThrProLysAspProGluIleLysAsnMetAsp 202
           ||| ::: || ::: :::
Qy   480 GAGAAA----------------ATGTGCCACCTGGTGTCGAGCTAACCAAGGG 518
           :::||| || :::--------GluAla 217
Db   203 GlnLysAspGlnAlaProGlnGluProGlnLysGlnPro---------------------

Qy   519 CTGGAGAGTGATGGAGCAGGAGGAGCCCACATGGAGAGTGACACAGACAGAGA 578
           ::: ||| |:::|||
Db   218 SerLysLysPro-GluThrGlnLysProProThrAsnProGluSerSerAsnThrGlnG1 237

Qy   579 CAGCGGCTATGGAGCAGGCTGAGGAGGAGGCCCGAGCAGGAGTGCAGGTGGCTGT 638
           ||| |||||||||:::
Db   237 nGluAsnLysGlnProGluValGlnLysThrProGluLysPro------------- 251

Qy   639 GGTCAGGATCAAGCCGCCCCTGCCCTCCCAG---GTAAACAGATTTACAGAGAAACTCAA 695
           || ||| ::: ||| ||
Db   252 ---------LysAlaProLysAlaProLysGluProGlnAsnProGlyAsnThrGlnAsnAs 268

Qy   696 CTGCAAAGCCCAACAAGAAATATAGCCCAGTGGGCAACTTCGAAAGGGAGATGGCACCAGTG 755
           ||| ||||| |||||||||||
Db   268 nAlaAsnAspGlnGlyGlnThrGlnGluAspSerAsnValGlnLysTrpGlnAsnse 288

Qy   756 GGCTGATGAACACATACAA------------------------TCCCAGAAGCTCAAT---CC 791
            :::||::: ||:::
Db   288 rGlyAlaHisGluValGluLeuThrLysIleSerSerAspGlyGlnAsnLeuAsnLeuse 308

Qy   792 TTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCTCACCAAAAGGAGATGA 851
           ||||||||||:::   :::
Db   308 rPheSerGluGluLeuProGlyIleThrAlaLysLeuValPheSerLysThrAspAs 328

Qy   852 GGGCTATGGCCCGCCCCAAAGAAGAAGACCAAAACTGCTGAAAGGGCCAAGCCGTGCTGAGGA 911
           ::: ||||| ::::::|||
Db   328 pLysSerGlyGlnThrLysGluIleAspPheSerThrThrLysSerLys 344

Qy   912 GCACATCTCAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATGGCTCGCCACAG 971
           ||:::
Db   344 --- 344

Qy   972 ACGAGATGGCAAGATCCAGGTTACTTTGGAGATTCTTTGACAGATACGTTCGTATTTC 1031
           :: ||| |||:::
Db   345 ---AspGlnLysIleAsnLeu-------------GlyGluLeuArgSerGlyLysTyrThrIleSe 361

Qy   1032 AGATAAAGTAGTGGGCATTCTCATG 1056
           ::: ||| |||:::
Db   361 rHisIleIlePheGlyIleAlaLeu 369

RESULT 13
Q9UQ39
ID   Q9UQ39   PRELIMINARY;   PRT;   956 AA.
AC   Q9UQ39;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA binding protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016088; BAA83714.1; -.
FT   NON_TER   1   1
FT   NON_TER   956   956
SQ   SEQUENCE   956 AA; 106462 MW; C9FEAD07B49F5011 CRC64;

Alignment Scores:
Pred. No.:   0.0397   Length: 956
Score: 132.00   Matches: 71
Percent Similarity: 36.50%   Conservative: 29

OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, partial cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016087; BAA83713.1; -.
FT   NON_TER   1   1
FT   NON_TER   1262   1262
SQ   SEQUENCE   1262 AA;   138003 MW;   75E76BD73FA14295 CRC64;

Alignment Scores:
Pred. No.:   0.0416
Score: 132.00
Percent Similarity: 36.50%   Length: 1262
Best Local Similarity: 25.91%   Matches: 71
Query Match: 6.41%   Conservative: 29
DB: 4   Mismatches: 88
                                           Indels: 86
                                           Gaps: 14

US-10-644-659A-1 (1-1146) x Q9UQ40 (1-1262)

Qy   6 TCCGGGCGAAAAGGAAAAGCGGGAGGGGCCC------------------AGCCAAGAG   44
          |||||||||||||||:::| ||| :::
Db   90 SerGlyArgArgGlyGluGlyAspAlaProPheSerGluProGlyThrThrSerThrGln 109

Qy   45 CGCCCT------CCGGAGATACGCACCAGCCACCTGGTCATCAGCTTGCCCAGGTTG   98
          ||||| ::::||||||||||
Db   110 ArgProSerSerProGluThrAlaThrLysGlnProSerSerProTyrGluAspLysAsp 129

Qy   99 GCAGCAGTGGGCGAATGAGAACAGCATCAGGAGGCCTACAGGCTGGCTGCC   158
          ||| ::: |||
Db   130 LysAspLysLysSerAlaThrArgProSerProSerProGluArgSerSerThr 149

Qy   159 GGGAGGGACCCA-----------------------------GGACTCACCTCA   182
          ||| :::||| |||||||||
Db   150 GlyProGluProAlaProThrProLeuLeuAlaGluArgHisGlyGly-SerProGl 169

Qy   183 A-------------------GCTCCTAAACCAATCACACCCCCTACT----TCACA   215
          | ::: :::|||||||:::|||
Db   169 nProLeuAlaThrProLeuSerGlnGluProValAsnProProSerGluAlaSerPr 189

Qy   216 CCAGAAAGCTCAGAGTGCCCCAAAGTCGCCACCCCGCCTGCCAGGAGGACATGGAGATGG   275
          ::: :::|||||||||||| :::|||||||
Db   189 oThrArgAspArgSerProLysSerProGluLysProGlyLeuProGlnSer---SerSerSe 208

Qy   276 ACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAG   318
          ::::||||||| ::: :::
Db   208 rGluSerSerProGlnProThrLysValSerArgHisAlaSerSerSerPr 228

Qy   319 ------------------------------GCTCCCACGCCGA-----GGAGAAAATGTGCCAA   341
          | ::: ||||||||
Db   228 oGluSerProLysProAlaProGlySerHisArgGluIleSerSerProTh 248

Qy   342 CAGCAAGACTTACGAGAGAGGA---------GGGGACGTGAGCCACCTCAGCCACCAGGTA   392
          ||||||| ::: |||
Db   248 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis-SerHisThrProS 268

Qy   393 CGAGAGGGATGTCGGTGTGCTTGAACCTGGCAGCCAGAGAATGACATTGACA-----445
          ::|||||| :::
Db   268 erArgArgMet----------------GlyArgSerArgSerProAlaThrAlaLysA 282

Qy   446 ----GAATCCTCCACAGCCACGGCTCCCAACGCGGA-------------GGAGAAATGTGCCAA   494
          |||||||| :::
Db   282 rgGlyArgArgSerArgSerArgThrProThrLysArgGlyHisSerArgSerArgSerProG 302

Qy   495 CCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGGAGG---AGAGCCCACATG   551
          ||||||||||| ||||||||||
Db   302 lnTrpArgArgSer-----ArgSerAlaGlnArgTrpGlyArgSerProGln- 319

Qy   552 GAGGAGTGACAGCGTAGACACAGAGACACAGAGGACAGCGCCTATGGAGGAGGAGGCC   611
          ||||||| ||||||||
Db   320 ArgArgGlyA   323

RESULT 14
Q9UQ40
ID   Q9UQ40   PRELIMINARY;   PRT;   1262 AA.
AC   Q9UQ40;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA binding protein (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Thu May 6 09:39:12 2004

us-10-644-659a-1.n2p.rspt

Page 12

```
Qy    612 CGAGCAGGATGGAGTGCAGGTGCTGCTGGTCAGGATCA 649
              |||||       |||         |||
Db    323 rgSerArgSerProGlnArgProGlyTrpSerArgSer 335

RESULT 15
Q9UQ35
ID   Q9UQ35           PRELIMINARY;         PRT;   2752 AA.
AC   Q9UQ35;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   RNA binding protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ohtaki S., Umeki K., Sawada Y.;
RT   "Homo sapiens mRNA for RNA binding protein, complete cds.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB016092; BAA83718.1; -.
DR   Genew; HGNC:16639; SRRM2.
SQ   SEQUENCE   2752 AA;  299672 MW;  109C64F181097123 CRC64;

Alignment Scores:
Pred. No.:               0.0475          Length:            2752
Score:                 132.00            Matches:             71
Percent Similarity:     36.50%           Conservative:        29
Best Local Similarity:  25.91%           Mismatches:          88
Query Match:             6.41%           Indels:              86
DB:                      4               Gaps:                14

US-10-644-659A-1 (1-1146) x Q9UQ35 (1-2752)

Qy      6 TCCGGCGAAAAGGAAAAGCGGGAGGGCCC---------------------AGCCAAGAG 44
            ||||||||||||:||||||:::                          ||    ::|||
Db    300 SerGlyArgArgGlyGluGlyAspAlaProPheSerGluProGlyThrThrSerThrGln 319

Qy     45 CGCCCT-------CCGGAAGATACGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTG 98
            ||||||       ||||||                |||
Db    320 ArgProSerProGluThrAlaThrLysGlnProSerSerProTyrGluAspLysAsp 339

Qy     99 GCAGCAGTGGACGGAATGAGAACAGCATCAGGCAGCGCCCAGGAGCCTACAGGCTGGCTGCC 158
            |||                              :::  |||        :::
Db    340 LysAspLysLysGluLysSerAlaThrArgProSerProSerProGluArgSerSerThr 359

Qy    159 GGGAGGGACCCA---------------------------------GGACTCACCTCA 182
            |||  ::|||||                                 ||| ||||||||
Db    360 GlyProGluProAlaProThrProLeuLeuAlaGluArgHisGlyGly-SerProGl 379

Qy    183 A--------------------GCTCCTAAACCAATCACACCCCCTACT-----TCACA 215
                                 :::  ||||||||||||||||| |||||||
Db    379 nProLeuAlaThrProLeuSerGlnGluProValAsnProProSerGluAlaSerPr 399

Qy    216 CCAGAAAGCTCAGAGTGCCCCAAAGTCGCCACCCCGCCTGCCAGAAGGACATGGAGATGG 275
            :::  ::|||||     ||||||||||||||||  :::|||||||
Db    399 oThrArgAspArgSerProLysSerProGlyLysLysProGlnSer---SerSerSe 418

Qy    276 ACAAAGCTCAGAGAAGCCCCTGAGTTTCTCACATCAAAAAG--------------------- 318
            ::|||||                          :::  ::: :::
Db    418 rGluSerProProSerProGlnProThrLysValSerArgHisAlaSerSerSerPr 438

Qy    319 -------------------------------AAAGAGGTGTCCAAAAACCGGTGGT 341
                                           :::|||||:::|||
Db    438 oGluSerProLysProAlaProAlaProGlyLysSerHisArgGluIleSerSerProTh 458

Qy    342 CAGCAAGACTTACGAGCAGAGA--------GGGGACGTGAGCCACCTCAGCCACAGGTA 392
            |||||||                       ||||||  |||  |||  |||
Db    458 rSerLysAsnArgSerHisGlyArgAlaLysArgAspLysSerHis-SerHisThrProS 478

Qy    393 CGAGAGGGATGCGTGGTGTGCTTGAACCTGGCCAGCAGAGAATGACATTGACA--------- 445
            ::|||                                            |||
```

Qy   478 erArgArgMet---------------GlyArgSerArgProAlaThrAlaLysA 492

Qy   446 -----GAATCTTCCACAGCCACGGCTCCCCAACGCGGA-------GGAGAAATGTGCCAA 494
               ||| ::: |||| ::||||||||

Db   492 rgGlyArgSerArgSerArgThrProThrLysArgGlyHisSerArgSerArgProG 512

Qy   495 CCTGGTGTCTGCTAACCAAGGGCTCGGAGAGTGATGGAGCAGG---AGGAGCCCACATG 551
            ||| ||||| ::: |||||||||

Db   512 lnTrpArgArgSer-------ArgSerAlaGlnArgArgTrpGlyArgSerArgSerProGln- 529

Qy   552 GAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCTGAGGAGAGGCC 611
            |||||||||||||||||||||||||||||||||||||||||||||||||||

Db   530 ------------------------------ArgArgGlyA 533

Qy   612 CGAGCAGGATGGAGTGCAGGTGCTGCTGGTCAGGATCA 649
            ||||| ||| |||

Db   533 rgSerArgSerProGlnArgProGlyTrpSerArgSer 545

Search completed: May 4, 2004, 08:54:52
Job time : 76 secs

```
OM nucleic - nucleic search, using sw model

               Copyright   GenCore version 5.1.6
               Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:   May  4, 2004, 23:55:43 ; Search time 2861 Seconds
                                   (without alignments)
                                   11961.572 Million cell updates/sec

Title:         US-10-644-659A-1
Perfect score: 1146
Sequence:      1 atggctccgggcgaaagga........tgattacgctaccaagtga 1146

Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched:      27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters:   55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    EST:*
              1: em_estba:*
              2: em_esthum:*
              3: em_estin:*
              4: em_estmu:*
              5: em_estcv:*
              6: em_estpl:*
              7: em_estro:*
              8: em_htc:*
              9: gb_est1:*
             10: gb_est2:*
             11: gb_htc:*
             12: gb_est3:*
             13: gb_est4:*
             14: gb_est5:*
             15: em_estfun:*
             16: em_estom:*
             17: em_gss_hum:*
             18: em_gss_inv:*
             19: em_gss_pln:*
             20: em_gss_vrt:*
             21: em_gss_fun:*
             22: em_gss_mam:*
             23: em_gss_mus:*
             24: em_gss_pro:*
             25: em_gss_rod:*
             26: em_gss_phg:*
             27: em_gss_vrl:*
             28: gb_gss1:*
             29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| | | % Query | | | | |
Result No.	Score	Match	Length	DB	ID	Description
1	792.8	69.2	1763	11	AK081696	AK081696 Mus muscu
2	696.8	60.8	2149	11	AK045186	AK045186 Mus muscu
3	618	53.9	618	29	AY399108	AY399108 Homo sapi
4	616.4	53.8	618	29	AY399109	AY399109 Pan trogl

No.	c	Score	%	Length	DB	ID	Accession
5		501.2	43.7	896	10	BF670302	BF670302 602149566
6		494.8	43.2	717	13	BX671002	BX671002
7		481.4	42.0	537	13	BX500732	BX500732 DKFZp779K
8		474	41.4	618	29	AY399110	AY399110 Mus muscu
9		429	37.4	576	9	AI605651	AI605651 ma51d03.y
10		424.6	37.1	909	13	BX667447	BX667447
11		423.8	37.0	684	13	BY730420	BY730420
12		418.2	36.5	624	13	BX675669	BX675669
13		416.6	36.4	553	10	BF549525	BF549525 UI-R-C1-1
14		401	35.0	630	13	BX671003	BX671003
15		392.6	34.3	528	13	BX511434	BX511434
16	c	379.6	33.1	582	14	CF368915	CF368915 853657 MA
17		377.8	33.0	646	10	BB643410	BB643410
18		363.2	31.7	532	14	CF367562	CF367562 842321 MA
19		358.8	31.3	613	9	AI510175	AI510175 mi97a10.y
20		346.8	30.3	498	13	BX668737	BX668737
21		323.6	28.2	552	9	AV599392	AV599392 AV599392
22		316.8	27.6	670	10	BB649319	BB649319
23		284.6	24.8	463	10	AW744257	AW744257 ur30a06.y
24		277.2	24.2	654	13	BU310812	BU310812 603541586
25		277.2	24.2	654	13	BU312228	BU312228 603544226
26		267.2	23.3	448	12	BG793104	BG793104 UTSW_H8E1
27	c	248.4	21.7	404	9	AI414584	AI414584 ma51d03.x
28	c	243	21.2	410	9	AV599391	AV599391 AV599391
29	c	236	20.6	344	14	W09785	W09785 ma51d03.r1
30		229.8	20.1	624	13	BX740493	BX740493
31		219.8	19.2	432	9	AA039033	AA039033 mi97a10.r
32	c	218.6	19.1	401	13	BQ211290	BQ211290 UI-R-Dy1-
33	c	211	18.4	974	29	CNS03MEA	AL250651 Tetraodon
34	c	198.2	17.3	388	9	AI172339	AI172339 EST218341
35		198	17.3	779	14	CA969408	CA969408 CclX06a22
36		193.4	16.9	552	13	BX869477	BX869477
37		193.4	16.9	695	13	BX869001	BX869001
38		193.4	16.9	728	13	BX869706	BX869706
39		193.4	16.9	744	13	BX871943	BX871943
40		193.4	16.9	756	13	BX868214	BX868214
41		184	16.1	401	14	W36976	W36976 mb82f10.r1
42		183	16.0	275	14	W18106	W18106 mb84f07.r1
43		176	15.4	367	13	BY340299	BY340299 BY340299
44	c	175	15.3	319	9	AI716006	AI716006 UI-R-Y0-a
45		171.2	14.9	359	13	BY341219	BY341219 BY341219

ALIGNMENTS

```
RESULT 1
AK081696
LOCUS       AK081696      1763 bp    mRNA    linear    HTC 20-SEP-2003
DEFINITION  Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
            library, clone:C130068012 product:weakly similar to F36F2.1 PROTEIN
            [Caenorhabditis elegans], full insert sequence.
ACCESSION   AK081696
VERSION     AK081696.1  GI:26349314
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636
REFERENCE   2
AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20489374
PUBMED      11042159
```

```
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
          Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
          Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
          Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
          Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
          Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
          Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED    11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
          FANTOM Consortium.
TITLE     Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
TITLE     Analysis of the mouse transcriptome based on functional annotation
          of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1763)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
          Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
          Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
          Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
          Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
          Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
          Nakamura,M., Nishi,K., Saito,R., Saito,H., Sakai,K., Sakazume,N.,
          Sano,H., Saaaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
          Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
          Muramatsu,M. and Hayashizaki,Y.
TITLE     Direct Submission
JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp;
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222;
          Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..1763
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:C130068O12"
                     /db_xref="MGI:2414716"
                     /db_xref="taxon:10090"
                     /clone="C130068O12"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="16 days embryo"
     CDS             54..1181
                     /note="unnamed protein product; putative
                     weakly similar to F36F2.1 PROTEIN [Caenorhabditis elegans]
                     (SPTR|O62234; evidence: FASTY, 42.3%ID, 84.5%length;
                     match=423)"
                     /codon_start=1
                     /protein_id="BAC38297.1"
                     /db_xref="GI:26349315"
```

```
/translation="MAPGERERAGPAKSALRKVRTATLVINLARGWQQWANENSTKQ
AQEPAGWLPGATHDLPNAPKEAGPYQHAPKTLSPKPDRDGEQHSEBATEVSHIKRKG
VTRTVVSKAYERGGDVNYLSHRYENDGGVSEAIQRLSIPQPENDIDRILLHDSPTRRRKCTNL
VSELTKGWKVMEQOEEPTWKSDSVDTEDSGYGGDMEERPEQDAAPVAPARIKRPLLSQA
NRYSETLNCKAHRKYSQVDNLKGRWQWADEHVQSQKLNPFSDEFDYDLAMSTRLKKG
DEGYGRPKEGSKTAERAKRABEHIYREIMELCFVIRTMARHRRDGKIQVTFGELFDRY
VRISDKVVGILMRARKHGLVHFEGEMLWQGRDDHVVITLLE"

ORIGIN

      Query Match          69.2%;   Score 792.8;   DB 11;   Length 1763;
      Best Local Similarity 81.7%;   Pred. No. 2.9e-191;
      Matches 936;   Conservative   0;   Mismatches 192;   Indels 18;   Gaps 1;
```

```
Qy     1 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGCCCAAGAGAGCGCCCTCCGGAAAGATA    60
           ||||||||  || || |||||||||| ||  ||| ||  || |||||||||||||
Db    54 ATGGCTCCAGGAGAGAAAGGAAAGGAGGCCGGGCCGGCCAAGAGTGCCCTCCGGAAGGTC   113

Qy    61 CGCACAGCCACCCTGTCATCAGCTTGGCCCAGGTTGGCCAGCAGTGGGCAGCAGTGAGAAC   120
           ||||||||  ||||||| ||||  ||||||||||||||||||||||||||  || ||||||
Db   114 CGCACAGCAACCCTGTTATCAATTTGGCCCAGGTTGGCCAGCAGTGGGCGGCGAATGAGAAC   173

Qy   121 AGCATCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCT        180
           || | |||  |||||| ||| |||||||||||||||||||  ||  || | |||  || |||
Db   174 AGTACCAAACAGGGCCCAGGACCTGCAGGCTGGCTGCCGGGAGCAACTCATGACCTACCT   233

Qy   181 CAAGTCCTAAACAATCACACCCCTACTTCACACAGAAAGCTCAGAGTGCCCAAAG         240
           || ||||||||  | ||     |||||    ||||||||||||||||||  ||||
Db   234 AACGCTCCTAAAGAAGCCGGTCC------------TTACCAGCATGCCCCCAA           275

Qy   241 TCGCCACCCGGCTGCCAGAAGGACATGGAGATGGAGAAAGCAAAGTTCAGAGAAGCCCCTGAG   300
           ||| || || |||||||| || |||  ||   | ||||  |  || |||||||||||| ||||
Db   276 ACTCGTCTCCAAGCCAATCGAGATCGAGAAGAGTCAGAGCAACAGTTCAGAGAAGAGCACCGAG   335

Qy   301 GTTTCCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAGAGA               360
           || ||||  |||||  |||||||| | |  |||||  || |||||||| |||||||
Db   336 GTCTCCCCAATTAAAAGGAAAGAGTGACCAGAGACGGTTGTCAGCAAGGCTTATGAGAGG   395

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACAGTACGAGAGGGGATGCTGGTGGTCTTGAACCT   420
           |||||  ||||||||||||||||||||||||||  ||||| ||| |||| | |||||| |||
Db   396 GGAGAGATGTGAACTACCTGAGCCACAGTTATGAGAATGAATGGCGGCGTCTCCCAACGCGGAGG   455

Qy   421 GGGCAGCAGGCAGAATGACATTGACAGAATCCTCCAGACGCTGAGG                 480
           ||||||||||| ||||||||||||||||| ||||  | | | || |||
Db   456 ATTCAGCCAGAATGACATTGACAGAATCCTTCTTAGTCACGACTCGCCAACACGGAGA      515

Qy   481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGAGGCAGGAG   540
           || |||||  |||||||||||| |||||  | |||||||||||| ||||| ||||||||  |
Db   516 AGAAAATGCACCAACCTGGTGTCTGAGCTCGACCAAAGGCTGGAAAGTGATGGAACAGGAA   575

Qy   541 GAGCCCCATGGAGGAGGTGACACCGCTAGACCAGGACACAGAGGAGGCT               600
           ||||| || ||| |||| ||  ||||| ||  || | ||||||   || |||||||
Db   576 GAGCCCACGTGGAAAGTGACACCGCTAGAGCCTAGGGCTACGGAGGGGATATG            635

Qy   601 GAGGAGGCCCGAGCCAGGATGGAGTGCAGGTGCCTGTGTCAGGATCAAGCGCCCCTTG      660
           ||||||||||||| || | || ||| | ||||  | |||||| ||||| |||  |||||
Db   636 GAGGAGGCCTGAGCCAGGCTGCAAGATGCGCCCTGGCTGCCTGCAGGATCAAACGCCCCTTG   695

Qy   661 CCCTCCCAGTGAAACAGATTTACAGAGAACTAGACTCAAGCCAACAGAAATATAGC        720
           ||| || |||| ||||| |||| |||||||||||  |||| |||  ||||||||
Db   696 CTCTCCCAGTGACACTCTGAACTGACAGCCCATCGGAATACAGC                   755

Qy   721 CCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACATACAATCCCAG     780
           || ||| ||||||||||||| |||  || |||| | |  |||||  || | | | || |||
Db   756 CAAGTGGACAACTTGAAAGGGGAGGTGGCCAGCAGTGGGCCGATGAACAACAGTTCCAGTCCCAG   815

Qy   781 AAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCCACCGCGCCTACAC          840
           |||||||||||||||||| |||  ||  || | |||| ||||  |||||
Db   816 AAGCTCAATCCTTCAGTGACGAATTTGACTATGATCTAGCCATGTCCACTCGGCTCCAC   875

Qy   841 AAAGGAGATGAGGGCTATGGCGCCCAAAGGAACCAAAACTGCTGAAGGGCCAAG        900
           |||| || | | ||||||||||| | | | || | |||   |  | |||||||||
Db   876 AAGGGAGGACGAGGGCTATGGCCGCCCCAAAGAGGGAAGCAAGACAGCTGAAAGGGCCAAG   935
```

```
Qy    901  CGTCTGAGGAGCACATTCACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG  960
                ||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db    936  CGAGCGGAGGAGCAGCATCTATCGGGAAATTATGGGACTGTGCTTTGTTATCCGCACAATG  995

Qy    961  GCTCGCCACGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTGACAGATAC  1020
                ||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db    996  GCTCGCCACGACGAGATGGCAAGATCCAGGTTACTTCGGAGAACTCTTTGATCGCTAT  1055

Qy   1021  GTTCGTATTTCAGATAAAGTAGTGTGGCCATTCTCATGCGTGCCAGGAAACATGGACTGGTA  1080
                |||||  |||||||||||||||||||||||||  ||||||||||||||||
Db   1056  GTTCGAATTTCTGATAAAGTCGTGGGCATCCTCATGCGTGCCAGGAAACACGGACTGGTG  1115

Qy   1081  GACTTTGAAGGAGGAGATGTGTATGGCAAGGCCGAGATGACCATGTTGTGATTACGGCTACTC  1140
                ||||||||||||||||||||  ||||||||||||||  ||||||  ||
Db   1116  CACTTTGAAGGAGGAGATGCTATGCCAAGGCCGAGACGACCATGTTGTGATTACTCTCCTT  1175

Qy   1141  AAGTGA  1146
           |||
Db   1176  GAGTAA  1181


RESULT 2
AK045186
LOCUS       AK045186                2149 bp    mRNA    linear    HTC 20-SEP-2003
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130044G13 product:weakly similar to
            F36F2.1 PROTEIN [Caenorhabditis elegans], full insert sequence.
ACCESSION   AK045186
VERSION     AK045186.1  GI:26637134
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing Pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
   MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2149)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,H., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
            Kanagawa Prefecture, Japan) whose assistance we gratefully
            acknowledge.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2149
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130044G13"
                     /db_xref="MGI:2410542"
                     /db_xref="taxon:10090"
                     /clone="B130044G13"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     CDS             <1..1009
                     /note="unnamed protein product; putative
                     weakly similar to F36F2.1 PROTEIN [Caenorhabditis elegans]
                     (SPTR|062234, evidence: FASTY, 42.3%ID, 84.5%length,
                     match=423)"
                     /codon_start=2
                     /protein_id="BAC32252.1"
                     /db_xref="GI:26337135"
                     /translation="STKQAQEPAGWLPGATHDLPNAPKEAGPYQHAPKTLSPKPDRDG
                     EGQHSEEATEVSHIKRKEVTRTVVSKAYERGGDVNYLSHRYENDGGVSEAIQPENDID
                     RILSHDGPTRRKCTNIVSELTKGWKVMEQEEPTWKSDSVDTEDSGYGGMEERPEQ
                     DAAPVAPARIKRPLLSQANRYSETLNCKAHRKYSQVDNLKGRWQQWADEHVQSKLNP
                     FSDEFDYDLAMSTRLHKGDEGYGRPKEGSKTAERAKRAEEHIYREIMELCFVIRTMAR
                     HRRDGKIQVTFGELFDRYVRISDKVVGILMRARKHGLVHFEGEMLWQGRDDHVVITLL
                     E"
ORIGIN
Query Match             60.8%;  Score 696.8;  DB 11;  Length 2149;
Best Local Similarity   81.0%;  Pred. No. 1.2e-166;
Matches  831;  Conservative   0;  Mismatches  177;  Indels  18;  Gaps   1;

Qy   121  AGCATCAGGCAGGCCAGGAGCCTACAGGCTGGCTGCCGGCGGAGGGACCCAGGACTCACCT  180
               || ||||||  ||||||||||||||||||||  |||||||||||||||||||||
Db     2  AGTACCAAACAGGCCCAGGAGCCTGCGGAGGCTGCTGCCCGGGAGCAACTCATGACCTACCT  61

Qy   181  CAAGCTCCTAAACAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAG  240
               |||||||||     ||||||| ||||||||||||||||||||     ||||||||
Db    62  AACGCTCCTAAAGAAGAAGCCGGTCCTTAC----------------CAGCATGCCCCCAAA  103
```

SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL   Science 302 (5652), 1960-1963 (2003)
PUBMED   14671302
REFERENCE   2 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Direct Submission
JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
     source   1..618
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     gene   <1..>618
                     /locus_tag="HCM0114"
ORIGIN

Query Match   53.9%; Score 618; DB 29; Length 618;
Best Local Similarity   100.0%; Pred. No. 7.7e-147;
Matches 618; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   529 ATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACAGAGGACAGCGCGGCTAT 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACAGAGGACAGCGCGGCTAT   60

Qy   589 GGAGGAGAGGCTGAGGAGAGGCCCGAGCAGGATGGAGTGGCAGGTGGCTGTGTGGTCAGGATC 648
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGAGGAGAGGCTGAGGAGAGGCCCGAGCAGGATGGAGTGGCAGGTGGCTGTGTGGTCAGGATC 120

Qy   649 AAGGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAA 708
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAGGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAA 180

Qy   709 CAGAAATATAGCCCAGTGGCAACTTGAAAGGGAGATGCAGCAGTGGCAGTGGCTGATGAACAC 768
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGAAATATAGCCCAGTGGCAACTTGAAAGGGAGATGCAGCAGTGGCAGTGGCTGATGAACAC 240

Qy   769 ATACAATCCCAGAGCTCAATCCTTCAGTGAGAGTTTGATTACGAGCTGGCCATGTCC 828
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATACAATCCCAGAGCTCAATCCTTCAGTGAGAGTTTGATTACGAGCTGGCCATGTCC 300

Qy   829 ACCCGCTCACACAAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCT 888
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACCCGCTCACACAAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCT 360

Qy   889 GAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATT 948
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATT 420

Qy   949 ATCTGCACAATGGCTCGCCACAGACGAGATGCAAGATCCAGGTTACTTTGGAGATCTC 1008
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATCTGCACAATGGCTCGCCACAGACGAGATGCAAGATCCAGGTTACTTTGGAGATCTC   480

Qy   1009 TTTGACAGATACGTTCGTATTCAGATAAGTAGTGGCATTCTCATGCGTGCCAGGAAA 1068
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTTGACAGATACGTTCGTATTCAGATAAGTAGTGGCATTCTCATGCGTGCCAGGAAA   540

Qy   1069 CATGGACTGGTAGACTTTGAAGGAGGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTG 1128

Qy   241 TCGCCACCCGCCTGCCGAAGGACATGGAGATGACAAAGCTTCAGAGAGAAAGCCCCTGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 ACTCTGTCTCCAAAGCCAGATCGAGACGGAGGGACACACTCAGAGAGAAGCCACCCAG 163

Qy   301 GTTTCTCACTCAACTCAAAAGAAGAGAGTGTCCAAAACGTGGTCCAAGACTTACGAGAGA 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 GTCTCCCACATTAAAGGAAAGAGAGTGACCAGAAACGGTTGTCAGCAAGCTTATGAGAGG 223

Qy   361 GGAGGGGACGTGAGCCACCTACCCACAGGTACGAGAGGGATGCTGGTGGTGCTTGAACCT 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 GGAGGAGATGTGAACTACCTGACCTGAGAATGATGCGGGCGTGTCAAGCTCTCGAAGCT 283

Qy   421 GGGACGCCAGAGAATGACATTGACAGAATCCTCCACAACGCGGAGG 480
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   284 ATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCACGAATCCGCCAACACGGAGA 343

Qy   481 AGAAAATGTGCCAACCTGGTTCTCGAGCTAACCAAGGGCTGAGAGTGAGAGTGACAGGAG 540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 AGAAAATGACCAACCTGGTTCTGAGCTGAGCTGACCAAAGGCTGGAAAGTGATGGAACAGGAA 403

Qy   541 GAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGCGGCTATGGAGGAGGCT 600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 GAGCCCACGGTGGAAGAGTGACAGCGTAGACACAGAGGACACAGAGGGGATATG 463

Qy   601 GAGGAGAGGCCCGAGCAGGATGGAGTGCGTGTGCTGCAGGTCAAGCGCCCCTTG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 GAGGAGAGGCCTGAGCAAAGTGCAGCGCGCCTGTGGCTCCTGCCAGGATCAAACGCCCCTG 523

Qy   661 CCCTCCCAGTGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCAACAGAAATATAGC 720
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 CTCTCCCAGGCAACAGAGTACTCTGAGACACTCAACTGTAAGGCCCATCGTAAGGCCGAAATACAGC 583

Qy   721 CCAGTGGCAACTTGAAAGGGAGATGGCAGCAGTGGCAGTGGCTCGATGATGGCCTGATGAACAATCCAG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 CAAGTGGCAACTTGAAAGGGAGGTGGCAGCAGTGGCAGCAGGGCCGATGATGAACACGTTCCAGTCCCAG 643

Qy   781 AAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGCCATGTCCACCCGCTACAC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 AAGCTCAATCCCTCCGTCAGTGACGAATTTGACTATGATCTTAGCACTATGGCCTTATCCACTCGGCTCCAC 703

Qy   841 AAAGGAGATGAGGGGCTATGCGGCGCCCCAAAAACTGCTGAAAGGGCCAAG 900
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 AAGGGGAAGCAGAGGGGTTATGCCGGCGCCCCAAAGAAGGCAGCCAAG 763

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAATGGCAATGTGCTTCATTATCTGCACAATG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   764 CGATCGGAAGAGCACATCTATCGGGAAATTATGGAACTGTGCTTGTTATCCGCACAATG 823

Qy   961 GCTCGCCACAGCGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   824 GCTCGCCACAGCGAGATGGGCAAGATCCAGGTTACTTTCCGGAGAACTCTTTGATCGCTAT 883

Qy   1021 GTTCGTATTTCAGATAAAGAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGACTGGTA 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   884 GTTCGAATTTTCTGATAAAGTCGTGGGCATCCTCATGCTGCCAGGAAACACCGGACTGGTCC 943

Qy   1081 GACTTTGAAGGAGGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTGATTACGCTACTC 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   944 CACTTTGAAGGAGGAGATGCTATGGCCAAGGCCGAGACGACCATGTTGTTGATTACTCTCCTT 1003

Qy   1141 AAGTGA 1146
           || |||
Db   1004 GAGTAA 1009

RESULT 3
AY399108

LOCUS   AY399108   618 bp   DNA   linear   GSS 12-DEC-2003
DEFINITION Homo sapiens HCM0114 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY399108
VERSION   AY399108.1 GI:39755097
KEYWORDS   GSS.

```
Db   541 CATGGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy  1129 ATTACGCTACTCAAGTGA 1146
         ||||||||||||||||||
Db   601 ATTACGCTACTCAAGTGA 618
```

RESULT 4
AY399109
LOCUS   AY399109   618 bp   DNA   linear   GSS 12-DEC-2003
DEFINITION Pan troglodytes HCM0114 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY399109
VERSION   AY399109.1 GI:37755098
KEYWORDS   GSS.
SOURCE   Pan troglodytes (chimpanzee)
  ORGANISM Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 618)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
  JOURNAL   Science 302 (5652), 1960-1963 (2003)
   PUBMED   14671302
REFERENCE   2 (bases 1 to 618)
  AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
     source   1..618
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
     gene   <1..>618
                     /locus_tag="HCM0114"
ORIGIN

Query Match   53.8%; Score 616.4; DB 29; Length 618;
Best Local Similarity   99.8%; Pred. No. 2e-146;
Matches 617; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   529 ATGGAGCAGGAGGAGGAGGAGCCACATGGAGGAGTGTTGTAGCAGCGTAGACACAGAGGACACAGCGGCTAT 588
         | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAGCAGGAGGAGGAGGAGCCACATGGAGGAGTGTTGTAGCAGCGTAGACACAGAGGACACAGCGGCTAT 60

Qy   589 GGAGGAGAGGCTGAGGAGGAGG 648
         ||||||||||||||||||||||
Db   61 GGAGGAGAGGCTGAGGAGGAGG 120

Qy   649 AAGCGCCCCTTGCCTCCCAGGTAAACAGATTACAGAGAAACTCAACTGACAAAGCCCAA 708
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAGCGCCCCTTGCCTCCCAGGTAAACAGATTACAGAGAAACTCAACTGACAAAGCCCAA 180

Qy   709 CAGAAATATAGCCCAGTGGGCAACTGAAAGGGAGATGCCAGCAGTGGGCTGATGAACAC 768
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGCCAGCAGTGGGCTGATGAACAC 240

Qy   769 ATACAATCCCAGAAGCTCAATCCTTTCAGTGAACAGTTTGATTACGAGCTGGCCATGTCC 828
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATACAATCCCAGAAGCTCAATCCTTTCAGTGAACAGTTTGATTACGAGCTGGCCATGTCC 300

Qy   829 ACCGGCCTACACAAAGGACAAAGGATGAGGAGGATGAGGAGCCGCCCCAAAGAAGAAGAACCAAAACTGCT 888
```

```
Db   301 ACCGGCCTACACAAAGGACAAAGGATGAGGAGGATGAGGAGCCGCCCCAAAGAAGAAGAACCAAAACTGCT 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   889 GAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGTACACATTGCTTCATT 948
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGTACACATTGCTTCATT 420

Qy   949 ATCTGCACAATGCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTGGAGATCTC 1008
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ATCCGCACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTGGAGATCTC 480

Qy  1009 TTTGACAGATACGTTCGTATTTCAGATAAAGTAGTAGTGGGCATTCTCATGCGTGCAGGAAA 1068
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTTGACAGATACGTTCGTATTTCAGATAAAGTAGTAGTGGGCATTCTCATGCGTGCAGGAAA 540

Qy  1069 CATGGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTG 1128
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CATGGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTG 600

Qy  1129 ATTACGCTACTCAAGTGA 1146
         ||||||||||||||||||
Db   601 ATTACGCTACTCAAGTGA 618
```

RESULT 5
BF670302
LOCUS   BF670302   896 bp   mRNA   linear   EST 21-DEC-2000
DEFINITION 602149566F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4290760 5',
            mRNA sequence.
ACCESSION   BF670302
VERSION   BF670302.1 GI:11944197
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 896)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institues of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1134 row: c column: 17
            High quality sequence stop: 525.
FEATURES   Location/Qualifiers
     source   1..896
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4290760"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_81"
                     /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                     (Clontech); Site_1: SfiI (ggcggcctcggcc); Site_2: SfiI
                     (ggccattatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C, G, or T). Average insert size
                     1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech Laboratories (Palo
                     Alto, CA)."
ORIGIN

Query Match   43.7%; Score 501.2; DB 10; Length 896;
```

RESULT 6

Best Local Similarity 98.9%; Pred. No. 7.2e-117;
Matches 526; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

```
Qy    618 GGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCTTGCCTCCCAGTAAACAG 677
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCTTGCCTCCCAGTAAACAG 60

Qy    678 ATTTACAGAGAAACTCAACTGCAAAGCCCAAGAAATATAGCCCAGTGGGCAACTTGAA 737
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAA 120

Qy    738 AGGGAGGATGGCAGCAGTGGGCTGATGAACACATACAATCCAGAAGTTCAATCCTTTCAG 797
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGGAGGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAGAAGTTCAATCCTTTCAG 180

Qy    798 TGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAAGGAGGATGAGGGCTA 857
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAAGGAGGATGAGGGCTA 240

Qy    858 TGGCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACAT 917
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACAT 300

Qy    918 CTACAGGGAAATGATGGACATGTTCATTATCTGCACAATGGCTCGCCACAGACGAGA 977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTACAGGGAAATGATGGACATGTTCATTATCTGCACAATGGCTCGCCACACGACGGA 360

Qy    978 TGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATACTTCGTATTTCAGAT-A 1036
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATACTTCGTATTTCAGATAA 420

Qy   1037 AAGTAGTGGCCATTCTCATGCGTGCCAGGAAACATGGACTTTGAGGAGGAGA 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAGTAGTGGCCATTCTCATGCGTGCCAGGAAACATGGACTTTGAGGAGGAGA 480

Qy   1097 TGCTATGGCAAGG--CCGAGATGACCATGTTGTGATTACGCTACTCAAGTGA 1146
          ||||||||||||    ||||||||||||||||||||||||||||||||||||
Db    481 TGCTATGGCAAGGGCCGAGATGACCATGTTGTGATTACGCTACTCAAGTGA 532
```

RESULT 6
LOCUS   BX671002   717 bp   mRNA   linear   EST 24-OCT-2003
DEFINITION BX671002 Sus Scrofa library (scac) Sus scrofa cDNA clone
            scac00351.h.22 3prim, mRNA sequence.
ACCESSION   BX671002
VERSION   BX671002.1 GI:37982096
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
  ORGANISM Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 717)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE   A Pig Normalised Multi-tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Clone distribution: AGENAE Resource centre. Francois PIUMI,
            Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
            genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
            FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at signasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0035 row: h column: 22.

FEATURES   Location/Qualifiers
     source   1..717
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scac00351.h.22"
                     /tissue_type="mixed"
                     /clone_lib="Sus Scrofa library (scac)"
                     /note="Vector: pT7T3D-pac vector; tissues: adipose tissue,
                     brain, kidney, liver, muscle, ovary, testis, heart,
                     hypothalamus, pancreas, skin, spleen, thymus, placenta,
                     pituitary gland, seminal vesicle, small intestine,
                     uterus, adrenals, bulbo uretral gland, cerebral trunk,
                     epididymis, female gonad, gall-bladder, hippocampus,
                     large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Query Match   43.2%; Score 494.8; DB 13; Length 717;
Best Local Similarity 89.6%; Pred. No. 2.7e-115;
Matches 532; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy    553 AGGAGTGACAGGTAGACACAGAGGACAGCGGCTATGGAGAGAGGGCTGAGGAGAGGGCCC 612
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGGAGTGATAGCATAGACACAGAGGACACAGCGGCTACGGAGGGGAGACAGAGGAGAGGGCCG 60

Qy    613 GAGCAGGAGGATGGAGGTGCAGGTGGCGTGTGTCAGGATCAAGGCGCCCTTGCCTCCCAGTA 672
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GAGCAGGAGGATGGAGGTGCAGGTGGCCATCGCCAGAATCAAACGCCCTTGCCCTCCCAGGCA 120

Qy    673 AACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAAC 732
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATAGATTTACCGAGAAACTCAATTGCAAAGCCCAGCGGAAATACAGCCAAGTGGGCCAC 180

Qy    733 TTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACAACAATCCCAGAAGAGTCAATCCT 792
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGAAAGGGAGGTGGCGCAGCAGTGGGCTGATGAACACATACAATCCCAGAAGAGTCAACCCT 240

Qy    793 TTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTGGCATGTGCATGTGCTGAGGAGGATGAG 852
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TTCAGTGACCAGTTTGATTATGAATTGGCCATGTGGCATGTGCATGTGCGACAAAGGAGATGAA 300

Qy    853 GGCTATGGCCGCCCCAAAGAAGAAGAACCAAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAG 912
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGCTACGGCCGTCCCAAGGAAGGAAGGAACCACGAAACTCGGGAAAAGGGCTAAGAGAGCTGAGGAG 360

Qy    913 CACATCTACAGGGAAAATGATGGACAATGTGCTTCATTATCTGCACATGGCTCGCCACACAGA 972
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CACATCTACAGAGAAAATGACAATGTGCTTCATCATCCGCACGATGCGCACCCCAGG 420

Qy    973 CGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATACTTCGTATTTCA 1032
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGGGACGGCAAGATCCAGGTTACTTTCGGAGATCTCTTTGACAGATACTTCGTATTTCA 480

Qy   1033 GATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTAGACTTTGAAGGA 1092
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATAAAGTAGTGGGCATTCTCATGCGCGGCAGGAAACACCGACTGGTTGACTTTGAGGGA 540

Qy   1093 GAGAGTGCTATGGCAAGGCCGAGATGGCCAAGGCCGAGATGACCACAGTTGTGATTACGCTACTCAAGTGA 1146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GAGATGCTATGGCAAGGCCGAGATGGCCAAGGCCGAGATGACCACCAGTTGTGATCGTCTGCTAAAGTGA 594
```

RESULT 7
LOCUS   BX500732   537 bp   mRNA   linear   EST 04-SEP-2003
DEFINITION DKFZp779K1758_r1 779 (synonym: hncc1) Homo sapiens cDNA clone
            DKFZp779K1758 5', mRNA sequence.
ACCESSION   BX500732
VERSION   BX500732.1 GI:32020385
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS   1 (bases 1 to 537)
             Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Obanger,A.,
             Fobo,G., Han,M. and Wiemann,S.
TITLE   EST (Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., et al.)
JOURNAL   Unpublished (2003)
COMMENT   Contact: MIPS
             MIPS
             Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
             This is the 5' sequence of the clone insert
             Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
             Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
             sequenced by BMFZ (Biomedical Research Center at the Heinrich-
             Heine-University, Duesseldorf/Germany) within the cDNA sequencing
             consortium of the German Genome Project. No s1 sequence available.
             This clone (DKFZp779K1758) is available at the RZPD in Berlin.
             Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
             Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES   Location/Qualifiers
     source 1..537
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="DKFZp779K1758"
             /tissue_type="liver"
             /dev_stage="fetal"
             /lab_host="DH10B"
             /clone_lib="779 (synonym: hnccl)"
             /note="Vector: pSport1_Sfi; Site_1: SfiIA; Site_2: SfiIB"

ORIGIN

Query Match   42.0%;   Score 481.4;   DB 13;   Length 537;
Best Local Similarity 99.8%;   Pred. No. 6.1e-112;
Matches   482; Conservative   0; Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 ATGGCTTCCGGCGAAAAGGAAAAGCGGGAGGCGGGGAGGGCCCTTCCGGAAGATA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55 ATGGCTTCCGGCGAAAAGGAAAAGCGGGAGGCGGGGAGGGCCCTTCCGGAAGATA 114

Qy   61 CGCACAGCCACCCTGGTCATCGCGTTGGCCCGGAGGTTGGCAGTGGGCGAATGAGAAC 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   115 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGGAGGTTGGCAGTGGGCGAATGAGAAC 174

Qy   121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGAGGACCCAGGACTCACCT 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   175 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGAGGACCCAGGACTCACCT 234

Qy   181 CAAGCTTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTTCAGAGTGCCCCAAAG 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   235 CAAGCTTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTTCAGAGTGCCCCAAAG 294

Qy   241 TCGCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGTCAGAGAAAGCCCCTGAG 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   295 TCGCCCACCCGCCTGCCAGAAGGACATGGAGATGGACAAAGTCAGAGAAAGCCCCTGAG 354

Qy   301 GTTTCTCACATCAAAAAGAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACAGAGAGA 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   355 GTTTCTCACATCAAAAAGAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 414

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCCAGGTACGAGAGGATGGCTGTGGCTTGAACCT 420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   415 GGAGGGGACGTGAGCCACCTCAGCCCAGGTACGAGAGGATGCTGTGGCTTGAACCT 474

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCCCACGCGCGAGG 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   475 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCCCACGCGCGAGG 534

Qy   481 AGA 483
            |||
Db   535 AGA 537

RESULT 8

AY399110
LOCUS   AY399110   618 bp   DNA   linear   GSS 12-DEC-2003
DEFINITION Mus musculus HCM0114 gene, VIRTUAL TRANSCRIPT, partial sequence,
            genomic survey sequence.
ACCESSION   AY399110
VERSION   AY399110.1 GI:39755099
KEYWORDS   GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Inferring nonneutral evolution from human-chimp-mouse orthologous
            gene trios
JOURNAL   Science 302 (5652), 1960-1963 (2003)
PUBMED   14671302
REFERENCE   2 (bases 1 to 618)
AUTHORS   Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
            Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
            Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
            Adams,M.D. and Cargill,M.
TITLE   Direct Submission
JOURNAL   Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
            Rockville, MD 20850, USA
COMMENT   This sequence was made by sequencing genomic exons and ordering
            them based on alignment.
FEATURES   Location/Qualifiers
     source 1..618
             /organism="Mus musculus"
             /mol_type="genomic DNA"
             /db_xref="taxon:10090"
     gene   <1..>618
             /locus_tag="HCM0114"

ORIGIN

Query Match   41.4%;   Score 474;   DB 29;   Length 618;
Best Local Similarity 85.4%;   Pred. No. 5.2e-110;
Matches   528; Conservative   0; Mismatches   90;   Indels   0;   Gaps   0;

Qy   529 ATGGAGCAGGAGGAGCCCACATGGAGGAGGAGTGCAGCGTAGACACAGAGGACAGCGGCTAT 588
            ||||||| | ||| |||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGGAACAGGAGAAGAGCCCAGCTGGAAGAGTGCACGCGTAGACACAGAGGACAGCAGTGGCTAC 60

Qy   589 GGAGGAGGAGGAGGAGGACCCGAGCCTGAGCCAAGATGCAAGAGATGAGTCGAGTGCGTGGTCAGGATC 648
            ||||||| | |||||||||||| |||||||||||||||||||||||||||||||
Db   61 GGAGGGGGGATATGGAGGAGGAGGACCCTGAGCCTGAGCAAGATGCAAGAGATGCAGCGCCTGTGCGCTGGTCCTGCCAGGATC 120

Qy   649 AAGCGGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGTAAACAGATTTACAGAGCTCAACTGCAAAGCCCAA 708
            || |||| |||| | || ||||| ||||||||||||||||||||||||||||||
Db   121 AAACGGCCCCTTGCTCTCCCAGGCAAACAGGTACTCTGAGCCAATGTCAACTGCAAGCTCAACTGTAAGGCCCAT 180

Qy   709 CAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCGTGATGAACAC 768
            ||| | |||||| ||||||| |||||||||||||||||||||||| ||| | | |||||||
Db   181 CGGAAATACAGCCAAGTGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCGTGGGCCGATGAACAC 240

Qy   769 ATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTTCGCCATGTCC 828
            || || |||||| || ||| |||||| ||||| | ||| |||| ||||| ||||||||||
Db   241 GTCCAGTCCCAGAAGCTCAATCCCTTCAGTCGTGACGAATTTGACTATGACCATGTCC 300

Qy   829 ACCCGCCTACACAAGGAGATGAGGCTATGGCCCGGCCCCCAAAGAAACCAAAACTGCT 888
            |||| |||||||||||||||| | ||||| ||| | |||||||||||||||| ||||
Db   301 ACTCGGTCCACAAGGGAGACAGGGGCTATCGCCGCCCAAAGAGAGGGAAGCAAGACAGCT 360

Qy   889 GAAAGGGCCAAGCGTGCTGAGGAGCACATCCACAGGGAAATGATGGAAATGGCTTCATT 948
            ||||||||||||| | ||| ||||| || ||| || ||||| ||| || ||| |||||
Db   361 GAAAGGGCCAAGCGAGCGGAGGAGCACAATCATCGGGAAATTATGGAAATTATGGCTTTTGTT 420

Qy   949 ATCTGCACAATGGCTTCGCCACAGACGAGATGCCAGGATCCAGGTTACTTTTGGAGATCTC 1008
            ||| ||| |||||| || ||| |||||||||

```
Db    421 ATCCGCACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTCGGAGAACTC 480
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1009 TTTGACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAA 1068
               |||||| ||||| |||||||||||||  |||||||||||||||||||| |||||||||||
Db    481 TTTGATCGCTATGTTCGCATTTCTGATAAAGTCGTGGGCATCCTCATGGGTGCCAGGAAA 540

Qy   1069 CATGGACTCGGTAGACTTTGAAGGAGAGAGATGCTATGGCAAGGCCGAAGATGACCATGTTGTG 1128
               ||| |||||||||| ||| ||||||||||||||||||||||||||||||||||||||||||
Db    541 CACGGACTGGTGCACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGACGACCACCATGTTGTG 600

Qy   1129 ATTACGCTACTCAAGTGA 1146
               |||| |||  |||
Db    601 ATTACTCTCGTTGAGTAA 618
```

```
RESULT 9
AI605651
LOCUS       AI605651                          576 bp     mRNA        linear     EST 15-MAR-2000
DEFINITION  ma51d03.y1 Soares mouse p3NMF19.5 Mus musculus cDNA clone
            IMAGE:314213 5' similar to WP:F36F2.1 CB15975 ;, mRNA sequence.
ACCESSION   AI605651
VERSION     AI605651.1  GI:4614818
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
            Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
            Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read is a RSEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            Possible reversed clone: similarity on wrong strand
            MGI:1204829
            Seq primer: -40RP from Gibco
            High quality sequence stop: 466
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..576
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:314213"
                     /dev_stage="19.5 dpc total fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares mouse p3NMF19.5"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAACGGCCGCCGCCATTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. RNA was kindly provided by
                     Dr. Minoru Ko (Wayne State University)."
ORIGIN
```

```
Query Match                    37.4%;  Score 429;  DB 9;  Length 576;
Best Local Similarity   83.9%;  Pred. No. 1.6e-98;
Matches 483;  Conservative   0;  Mismatches  93;  Indels   0;  Gaps   0;

Qy    406 GGTGTGCTTGAACCTGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGC 465
               | ||||||||||||  |||| |||||||||| |||||||||||||||  |||||||
Db      1 GGCGTGTCTGAAGCTATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCACGAC 60

Qy    466 TCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGGAGA 525
               ||| |||||||||||||||| ||| ||| ||||||| |||||||||||||||||||
Db     61 TCGCCAACACGGAGAGAGAAATGCACCAACCTGGTCTGACGCTGACCAAGGGCTGGAAA 120

Qy    526 GTGATGGAGCAGAGGAGAGCCCATGAGGAGTGAGCGGTGAGACAGAGAGACAGCGGC 585
               |||||| ||||||||||||| ||||| ||||||||| ||||||||||||||||||||
Db    121 GTGATGGAACAGGAGAAGAGCCCACGTGGAGAGGTGACAGCGTAGACACAGAGAGCAGTGGC 180

Qy    586 TATGGAGAGAGGCTGAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTTGGCTGTGGTCAGG 645
               |||||||| |||||||||||||||||||||||||| ||||| ||||||||| |||||||
Db    181 TACGGAGGGGATATGGGAGGAGGAGGCCTGGACCAAGATGCAGCGGCCTGTTGGCTCCTGCCAGG 240

Qy    646 ATCAAGGGCCCCTTGCCCTCCCAGGTAAACAGAATTTACAGAGAAACTCAACTGCAAAGCC 705
               ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATCAAACGCCCCTTTGCTCTCCCAGGCAAACAGGTACTCTGAGACACTCAACTGTAAGGCC 300

Qy    706 CAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGAGGATGGCAGCAGTGGGCTGATGAA 765
               ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    301 CATCGGAAATAACAGCCAAGTGACAACTTGAAAGGAGGTGGCAGCAGTGGGCTGATGAA 360

Qy    766 CACATACAATCCCAGAGCTCAATCCTTTCAGTGAAGAGTTTGATTAGCGAGTGTGGCCATG 825
               |||||| |||| ||||||||||||||||||||||||||||||||| ||| ||||||||||
Db    361 CACGTCCAGTCCCAGAGCTCAATCCTTCAGTGACGAATTTGACTATGATCTAGCCATG 420

Qy    826 TCCACCCGCTACACAAAGGAGGATGAGGGCTATGGCCCCAAAGAAGGAACCAAAACT 885
               ||||| ||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    421 TCCACTCCGCTCCACAAGGGAGACGACGAGGGCTATGGCCCCAAAGAAGGAGGAAGCAAGACA 480

Qy    886 GCTGAAGGGCCAAGCGTGCTGAGGAGCACCATCTACAGGGAAATGATGGACATGTGCTTC 945
               ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db    481 GCNTANAGGGCCAAGCGAGCGGAGAGCACCATCTATCGGAAATTATGGAACTGTGCTNT 540

Qy    946 ATTATCTGCACAATGGCCTGCGCCACAGACGAGATGGC 981
               |||| |||||||||||||||||||||||||||||||||
Db    541 GTTATCCGCCACAATGGCTCGCCACAGACGAGATGGC 576
```

```
RESULT 10
BX667447
LOCUS       BX667447                          909 bp     mRNA        linear     EST 22-OCT-2003
DEFINITION  BX667447 Sus Scrofa library (scag) Sus scrofa cDNA clone
            scag0001c.b.12 3prim, mRNA sequence.
ACCESSION   BX667447
VERSION     BX667447.1  GI:37857165
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 909)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Bonaldo,F. and Hatey,F.
  TITLE     A Pig Normalised Multi-Tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Clone distribution: AGENAE Resource centre, Francois PIUMI,
            Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
            genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
```

FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73
Sequence cleaned of vector, adaptor and repetitions. Contact us
at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
sequence.

FEATURES   Plate: 0001 row: b column: 12.
    source   Location/Qualifiers
                     1..909
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scag0001c.b.12"
                     /tissue_type="mixed"
                     /dev_stage="from embryos to adults"
                     /clone_lib="Sus Scrofa library (scag)"
                     /note="vector: pT7T3D-pac vector; tissues: adipose tissue,
                     brain, kidney, liver, muscle, ovary, testis, heart,
                     hypothalamus, pancreas, skin, spleen, thymus, placenta,
                     pituitary gland, seminal vesicle, small intestine,
                     uterus, adrenals, bulbo uretral gland, cerebral trunk,
                     epididymis, female gonad, gall-bladder, hippocampus,
                     large intestine, male gonad, melanocytes, stomach, udder"

ORIGIN

Query Match   37.1%;   Score 424.6; DB 13; Length 909;
Best Local Similarity 80.5%;   Pred. No. 2.8e-97;
Matches 598; Conservative 0; Mismatches 120; Indels 25; Gaps   8;

Qy   403 GCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCAGCCAC 462
          || || ||||| ||||||||| || ||||| ||||||||| |||||| |||||||||||
Db   1 GGTTGATGAGCCTGANCCCGAGCAGCCAGAGAGTGACATGNACAGACTCCTCCGCAGCCAT 60

Qy   463 GGCTCCCAACGCGGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGGCTGG 522
          |||| ||||| ||||| ||||| ||||| ||||||||||||||| ||||| ||||||||||
Db   61 GGCTCCCAACGCGCGAGAGAAAATGTGCCAACCTGGTATCTGAGCTGGACCAAGGGGCTGG 120

Qy   523 AGAGTGAT------GGAGCAGGAGGAGCCCACATGGAGGAGTGACGCGTAGACACA 573
          ||||| || |||||||| ||||||||||| ||||||||| ||||| |||||||
Db   121 AAGAGATGGAACAGGAGGAGGATCAGGAGGCTCAAATGCAGGAGTGATAGCATAGACACA 180

Qy   574 GAGGACAGCGGCGTATGGAGGAGAGGCTGAGGAGGAGCCCGAGCAGCAGGAGTGCAGGTG 633
          || ||||||||| ||||||||| || ||||| ||||| ||||||||||||||||||||||
Db   181 GAAGACAGCGGCGTACGGAGGAGACGAGGAGGAGCCCGAGCAGCAGGATGAGAGCAGGTG 240

Qy   634 GCTGTGGTCAGGATCAAGCGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTC 693
          || ||||| |||||||||| ||||||| ||||||||| || ||||| ||||||||||||||
Db   241 GCCATCGCCAGAATCAAACGCCCCTTGCCCTCCCAGGCAAATAGATTTACCGAGAAACTC 300

Qy   694 AACTGCAAAGCCCACAGAAATATAGCCCCAGTGGGCAACTTGAAAGGGAGATGCAGCAG 753
          ||||| ||||| |||| ||||||| || ||||||||| ||||| ||||||||||||||||
Db   301 AATTGCAAAGCCCAGCGAAATACAGCCAAGTGGGCACGTGGGCAGCCACCTGGCCGCAGCAG 360

Qy   754 TGGGCTGATGAACACATACAATCCCAGAGAGCTCAATCCTTTCAGTGAAGAGTTTGATTAC 813
          |||||||||||||||||| ||||||||||||||||| ||||| |||||||||||||||||
Db   361 TGGGCTGATGAACACATACAATCCCAGAGAGCTCAACCCTTTCAGTGACGAGTTTGATTAT 420

Qy   814 GAGCTGGCCATGTCCACCCGCCTACACAAGGAGAGTGAGGCGTATGGCCGCCCCAAAGAA 873
          ||||||||||||| ||||| || || |||||||| || ||| ||| ||||||| |||||
Db   421 GAATTGGCCATGTCCACCCGCCTGCACAAGGAGATGAGGCTACGGCCGTCCAAGGAA 480

Qy   874 GGAACCAAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGG-GAAATGAT 932
          ||||||||| ||||||||||||||| ||||||||||||||| ||||||||||| |||||||
Db   481 GGAACCAGAACTGCGGAAAGGGCTTAAGAGAGCTGAGGAGGCCATCCACAGAGAAATCAT 540

Qy   933 GGACATGTGCTTCATTATCTGCACA---ATGGCTCGCCACAGACGAGATGGCAAGATCC 988
          |||||| ||||| ||||| ||||||| |||||||| ||||| ||||| |||||||||||
Db   541 GGACATGTGCTTCTATCATCNCGCACGCGCCACNCCCAGGGCGGAGCGCGGCCAAGATCN 600

Qy   989 AGGTTACTTTTGGAGA-----TCTCTTTGACAGATACGTTCGTATTCGAGATA-AAGTAG 1042
          ||||||||| |||||| ||| |||||||| ||||| ||||||||||| |||| ||||
Db   601 CAGGTTACTTTCGGGAGATCTTCTTTGACAGATACGNTTCGTATTCGAGATANAAGTAG 660

Qy   1043 TGGGCATTCTC---ATGGCTGCCAGGAAACATGGAACTGTAGACTTTGAAGGGAG-ATG 1098
          ||||||| ||||||||||||| || ||| |||

genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.

FEATURES
    source   Location/Qualifiers
                    1..684
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /db_xref="taxon:10090"
                    /clone="D330042E04"
                    /tissue_type="heart"
                    /dev_stage="13 days embryo"
                    /lab_host="DH10B"
                    /clone_lib="RIKEN full-length enriched, 13 days embryo
                    heart"
                    /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                    prepared and sequenced in Mouse Genome Encyclopedia
                    Project of Genome Exploration Research Group in Riken
                    Genomic Sciences Center and Genome Science Laboratory in
                    RIKEN. Division of Experimental Animal Research in Riken
                    contributed to prepare mouse tissues. 1st strand cDNA was
                    primed with a primer [5'
                    GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was
                    prepared by using trehalose thermo-activated reverse
                    transcriptase and subsequently enriched for full-length by
                    cap-trapper. Second strand cDNA was prepared with the
                    primer adapter of sequence [5'
                    GAGAGAGAGATTCTGAGTTAATTAAATTAAATCCCCCCCCCCCCC 3']. cDNA
                    was cleaved with BamHI and XhoI. Vector: a modified
                    pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Query Match   37.0%;   Score 423.8;   DB 13;   Length 684;
Best Local Similarity 77.6%;   Pred. No. 3.8e-97;
Matches 545;   Conservative   0;   Mismatches 138;   Indels 19;   Gaps 2;

Qy   121 AGCATCAGGCAGGCAGGCCTACAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCT 180
             ||||||||| |||||| ||||| || |||||| ||||||| | ||||| ||||
Db   2 AGTACCAAACAGCCCCAGGAGCCTGCGAGCGCCTGGCTGCCGGGGAGCAACTCATGACCTACCT 61

Qy   181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACACCAGGAAAGTCAGAGTGCCCCAAAG 240
             |||||||| ||||||| ||||| ||||||| || |||
Db   62 AACGCTCCTAAAGAGAGCCGGTCCTTAC------------CAGCATGCCCCCAAA 103

Qy   241 TCGCCACCCGGCCTGCCAGAAGGACATGGACAAGGACGGAGAGGGACCCCTGAG 300
             ||||||||||||||||||| ||| ||||||||| || || |||||||||||
Db   104 ACTCTGTCTCCAAAGCCAGATCGAGACGCGAGAGGGACCAACTCAGAAGAGCCACCGAG 163

Qy   301 GTTTCTCACATCAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
             | ||||| || |||| |||| || ||| ||| |||| ||||||| ||||||||
Db   164 GTCTCCCACATTAAAAGGAAAGAGGTGACCAGAACGGTTGTCAGCAAGGCTTATGAGAGG 223

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACCAGGTACGAGAGGGATGCTCGTGTCTTGAACCT 420
             || |||||| |||| |||| |||| ||||| ||||||| |||| ||| ||||| |||
Db   224 GGAGGAGATGTGAACTACCTGACCACCACAGGTATGAGAATGATGGGCGCGTGTCGAAGCT 283

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCGTCCCCAAACGGGAGG 480
             ||||||||||||||||| |||||||||| || | ||||| | | |||||| |||||
Db   284 ATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCACGACTCGCCACACCGGAGA 343

Qy   481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGCTGGTGAGTGAGCTGAGCAGGAG 540
             |||||||| |||||||||| ||||||| || ||||||| | || |||| ||||||||
Db   344 AGAAAATGCACCAACCTGGTGTCTGAGCTGACCAAGGCTGACCAAAGGCTGAAAGTGATGGAACAGGAA 403

Qy   541 GAGCCCACATGGAGTGAGGAGTGACAGCGGTGAACACAGAGGACAGCGGCTATGGAGGAGGGCT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   404 GAGCCCACGTGAAGAGTGACAGCGGTGAGACAGCAGAGACAGAGGACACGAGTGGCTACGGAGGGGATATG 463

Qy   601 GAGGAGAGGCCCGGACGAGGATGGAGTGCAGGTGCAGTTGGCTGTGGTCAGGATCAAGCGGCCCTTG 660
             ||||||||||||||||||| || |||||||||||||||||||||||||
Db   464 GAGGAGAGGCCTGGACGATGCAGCCGCCTGGGCTGGGCTCCTGCGGGATCAAACGCCCCCTTG 523

Qy   661 CCCTCCCAGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCAACAGAAATATAGC 720
             |||||||||||||||||||||||||||||||||||||||||
Db   524 CTCTCCCAGGCAAACAGGTACTCTGAGACACTCAACTGTAAGGCCCATC-GGAATACAGC 582

Qy   721 CCAGTGGCAACTTGAAAGGAGATGGGCAGCAGTGGGCTGATGAACACATACAATCCCAG 780
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db   583 CAAGTGGACAACTTGAAAGGGAGGAGGTGGCAGCCAGTGGGCCGATGAACACGTNCAGTCCCAG 642

Qy   781 AAGCTCAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCC 822
             ||||||||||| |||||||||||||||||||||||||||||
Db   643 AAGTCAATCCCTTCAGTGACGAATTTGACTATGATCTTAGCC 684

RESULT 12

LOCUS   BX675669   624 bp   mRNA   linear   EST 28-OCT-2003
DEFINITION BX675669 Sus Scrofa library (scac) Sus scrofa cDNA clone
            scac00301.n.16 3prim, mRNA sequence.
ACCESSION   BX675669
VERSION   BX675669.1   GI:38009804
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
  ORGANISM Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 624)
  AUTHORS   Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,
            Soares,M., Boraldo,F. and Hatey,F.
  TITLE   A Pig Normalised Multi-tissue cDNA Library
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Tosser-Klopp G
            Genetique Animale
            Institut National de la Recherche Agronomique
            Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan
            cedex, FRANCE
            Tel: 33 (0) 5.61.28.51.14
            Fax: 33 (0) 5.61.28.53.08
            Email: tosser@toulouse.inra.fr
            Clone distribution: AGENAE Resource centre. Francois PIUMI,
            Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et Etude du
            genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,
            FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at sigenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0030, row: n column: 16.

FEATURES   Location/Qualifiers
    source   1..624
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="scac00301.n.16"
                     /tissue_type="mixed"
                     /clone_lib="Sus Scrofa library (scac)"
                     /note=Vector: pT7T3D-pac vector; tissues: adipose tissue,
                     brain, kidney, liver, muscle, ovary, testis, heart,
                     hypothalamus, pancreas, skin, spleen, thymus, placenta,
                     pituitary gland, seminal vesicle, small intestine,
                     uterus, adrenals, bulbo uretral gland, cerebral trunk,
                     epididymis, female gonad, gall-bladder, hippocampus,
                     large intestine, male gonad, metanocytes, stomach, udder"

ORIGIN

Query Match   36.5%;   Score 418.2;   DB 13;   Length 624;

Best Local Similarity 90.3%; Pred. No. 9.7e-96; Indels 0; Gaps 0;
Matches 447; Conservative 0; Mismatches 48;

```
Qy    652 CGCCCCTTGCCTCCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAG 711
          ||||||||| |||||||||||| |||| |||||||||||||||||||||||||||||||
Db      1 CGCCCCTTGCCTCCCCAGGCAAATAGATTTACCGAGAAACTCAATTGCAAAGCCCAGCGG 60

Qy    712 AAATATAGCCCAGTGGCAACTTGAAAGGGATGGCAGCAGTGGCTGATGAACACATA 771
          ||||| |||| |||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAATACAGCCCAAGTGGCCACCTGAAAGGGAGGGTGGCCAGCAGTGGCTGATGAACACATA 120

Qy    772 CAATCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACC 831
          ||||||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||
Db    121 CAATCCAGAAGCTCAACCCTTTCAGTGACGACGAGTTTGATTATGAGGCCATGTCCACC 180

Qy    832 CGCCTACACAAAGGAGATGAGGGCTATGGCGCCCCAAAGAAGGAACCAAAACTGCTGAA 891
          ||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    181 CGCCTGCACAAAGGAGATGAGGGCTACGGCGCCGTCCCAAGGAAGGAACCAAACTGCTGGAA 240

Qy    892 AGGGCCAAGCCTGCTCGAGGAGCACATCTACAGGGAAAATGATGGACATGTGCTTCATTATC 951
          ||||| |||| |||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    241 AGGGCTAAGAGAGCTGAGGAGCACATCTACAGAGAAATCATGGACATGTGCTTCATCATC 300

Qy    952 TGCACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTT 1011
          |||||||||||||| || |||||||||||||||||||||||||||||||||||||||||
Db    301 CGCACGATGGCGCACCCCAGGCGGGACGCGGGAAGATCCAGGTTACTTTCCGGAGATCTCTT 360

Qy   1012 GACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCATCGGTGCCAGGAAACAT 1071
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    361 GACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCATCATGCGCGCCAGGAAACAC 420

Qy   1072 GGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATT 1131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 GGACTGGTTGACTTTGACTTTGAGGGAGAGATGCTATGGCAAGGCCCAGATCGACCACGTTGTGATC 480

Qy   1132 ACGGTACTCAAGTGA 1146
          |||||||||||||
Db    481 ACTCTGCTAAAGTGA 495
```

RESULT 13
BF549525

LOCUS   BF549525   553 bp   mRNA   linear   EST 12-DEC-2000
DEFINITION UI-R-C1-li-h-06-0-UI.r1 UI-R-C1 Rattus norvegicus cDNA clone
            UI-R-C1-li-h-06-0-UI 5', mRNA sequence.
ACCESSION   BF549525
VERSION   BF549525.1 GI:11659213
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 553)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is also available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1775884
            Seq primer: M13 Forward.
FEATURES   Location/Qualifiers

source   1..553
                /organism="Rattus norvegicus"
                /mol_type="mRNA"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-C1-li-h-06-0-UI"
                /dev_stage="Adult"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="UI-R-C1"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C1
                library is a subtracted library derived from the UI-R-C0
                library, which is a subtracted library derived from the
                UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library
                consisted of a mixture of individually tagged normalized
                libraries constructed from rat placenta, adult lung,
                brain, liver, kidney, heart, spleen, ovary, and muscle.
                The UI-R-E1 library consisted of a mixture of
                individually tagged normalized libraries constructed from
                8, 12 and 18-day embryo. The tag is a string of 3-5
                nucleotides present between the Not I site and the
                oligo-dT track which allows identification of the library
                of origin of a clone within the mixture. The subtracted
                library (UI-R-C1) was constructed as follows: PCR
                amplified cDNA inserts from UI-R-C0 clones from which 3'
                ESTs had been derived was used as a driver in a
                hybridization with the UI-R-C0 library in the form of
                single-stranded circles. The remaining single-stranded
                circles (subtracted library) was purified by
                hydroxyapatite column chromatography, converted to
                double-stranded circles and electroporated into DH10B
                bacteria (Life Technologies) to generate the UI-R-C1
                library. This procedure has been previously described
                (Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
                1996)"

ORIGIN

Query Match   36.4%;   Score 416.6; DB 10; Length 553;
Best Local Similarity 85.3%; Pred. No. 2.3e-95;
Matches 464; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

```
Qy    543 GCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGACGAGAGGCTGA 602
          |||||||||||||||||||| ||||||||||||||||| |||||||||||||||||||
Db      7 GCCCAAGTGGAAGAGTGACAGCGTAGCACAGAGGACAGTGGCTACGGACGGGGACCATGGA 66

Qy    603 GGAGAGGCCCGAGCGGATGGAGTGTCAGGTGGCTGTGGTCAGGATCAAGGCCCCTTGCC 662
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db     67 GGAGAGGCCCTGACGAAGATGTAGCGCGAGATGGTGGCCTCTGCCAGGATTAAACGCCCCTTGCA 126

Qy    663 CTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACACAGAAATATAGCCC 722
          ||||||||||||||||||||| ||||||||||||||||||||||||||||||  ||||||||
Db    127 CTCCCAGGCAAACAGATACTCAGGACACTCAACTGTAAGGCCATCGGAAATACAGCCA 186

Qy    723 AGTGGCAACTTGAAAGGGATGGCAGCAGTGGCTGATGAACACATACAATCCCAGAA 782
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187 AGTGGACAACTTGAAAGGGACGAGGTGGCCGAGCAGTGGCCGATGAACACATACAGTCACAGAA 246

Qy    783 GCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAA 842
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    247 GCTCAACCCTTTCAGTGATGAATTTGACATATGACCTAGCCATGTCCACTCGACTCGAGGAA 306

Qy    843 AGGAGATGAGGGCTATGGCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCG 902
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    307 GGGAGACGAGGGGATATGGCCCCCCCAAAGAAGGAGCCAAGACAGCTGAAAGGGCCAAGAG 366

Qy    903 TGCTGAGGAGCACATCTACAGGGAAAATGATGGACATGTGCTTCATTATCTGCACAATGGC 962
          |||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||
Db    367 AGCCGAGGAGCACATCATCGGAAATATGGAATTATGGTGTTTGTTGTTATCCCGCACAATGGC 426

Qy    963 TCGCCCAGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCCTTCTTGAGATCTTTGATAGATAGT 1022
          |||| |||||||||||||||||||||||||||||| ||||||| |||||||||||
Db    427 TCGCCATAGACGAGATGGCAAGATCCAGGTTACTTTCGGGAACTCTTTCGGGAACTCTTTGATAGATATGT 486
```

```
Qy   1023 TCGTATTCAGATAAAGTAGTGGGCATTCTCATGCTGCTGCCAGGAAAACATGGACTGGTAGA 1082          Db   157 TAGCACCAGGCAGGCCCAGGAGCCTACAGGCTGGATGCCAGGAGGGGCCCAGGGAGTCAGA 216
          ||| |||||||||||||||| |||||||| ||||||||||||||||||||| |||||||||               ||| ||||||||||| |||||||||||| |||||||| |||||||||||| ||||||||
Db    487 TCGAATTCAGATAAAGTGGTGGGCATCCTCATGCTGCTGCCAGGAAACACNGACTGGTGCA 546          Qy   180 TCAAGCTCCTAAACAATCACACCCCTACTTCACCACCAGGAAAGCTGCAGAGTGCCCCAAA 239
                                                                                           ||||| |||||| ||||||||||||||||||||||||||||||||||| |||||||||||
Qy   1083 CTTT 1086                                                                     Db    217 CCAACTTTCTGACCAGTGATACACCCCTACAACCACCACCAGAAAGCTCCAGAGTGCCCCAAA 276
          ||||                                                                                 |||||||| ||||||||| ||||||||||||||||||||||| |||||||||||||
Db    547 CTTT 550                                                                     Qy   240 GTCGCCACCCCGCCTGCCAGAAGGACATGGAAGCTGAGATGGACAAAGCTCAGAGAAAGCCCTGA 299
                                                                                           |||||||||| |||||||||||| |||||||| |||||| |||||||||||| |||||||
RESULT 14                                                                          Db    277 ATCTCCCTCCCAAGCCAGGGGATATGGAGCTGAGCTGGACAAAGCTCAGAGAGAGCCACCGA 336
BX671003                                                                            Qy   300 GGTTTCTCACATCAAAAGAAAGAGAGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAG 359
LOCUS        BX671003              630 bp    mRNA      linear   EST 24-OCT-2003              |||| |||||||||||||||| ||||||||||||| ||||||||||| ||||||||||||
DEFINITION   BX671003 Sus Scrofa library (scac) Sus scrofa cDNA clone                 Db    337 AGTCTCCTCCTATCAAAAGGAAAGAGAGTGACCAAGCACCATTGTCAGCAAAGCTTACAGAGAG 396
             scac0035i.h.22 5prim, mRNA sequence.                                           Qy   360 AGGAGGGGGACGTGAGCCACCTCAGCCACCAGGTACGAGAGGGGATGTCGTGTGTCTTGAACC 419
ACCESSION    BX671003                                                                          |||||||||| |||||| ||||||||||||||| |||||||||||||| |||||||||||
VERSION      BX671003.1   GI:3798098                                                  Db    397 AGGAGGTGATGTAAGCCACCTCAGCCATATACGAGATACGAGAAGGGTGTGATGAGCCTGAACC 456
KEYWORDS     EST.                                                                     Qy   420 TGGGCAGCCAGAAGTGACATTGACAGAATCCTCCACAGACGGCTCCCAACGCGCGGAG 479
SOURCE       Sus scrofa (pig)                                                                  ||| |||||||||||||||| ||||||||| ||||||||| ||||||||||| |||||||
ORGANISM     Sus scrofa                                                               Db    457 CGAGCAGCCAGAGGTGACATTGACAGACTCCTCCGCAGCCATGGCTCCCCAACGCGCGGAG 516
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;        Qy   480 GAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGAT--------- 530
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.                         ||||||||||||||||||||||||| |||||| ||||||| |||||||||||
REFERENCE    1 (bases 1 to 630)                                                       Db    517 AAGAAAATGTGCCAACCTGGTATCTGAGCTGACCAANGGGCTGGAAGGAGATGGAACAGGA 576
AUTHORS      Bonnet,A., Tosser-Klopp,G., Benne,F., Cabau,C., Villeger,S.,             Qy   531 GGAGCAGGAGGAGGCCCACATGGAGGAGTGACAGCGTGAGACACAGAGGACAGCGG 584
             Soares,M., Bonaldo,F. and Hatey,F.                                               ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
TITLE        A Pig Normalised Multi-Tissue cDNA Library                              Db    577 GGACACAGGAGGAGCTCAAATGCAGGAGTGATAGCATAGACACAGAAGACACAGCGG 630
JOURNAL      Unpublished (2003)
COMMENT      Contact: Tosser-Klopp G
             Genetique Animale                                                       RESULT 15
             Institut National de la Recherche Agronomique                            BX511434
             Chemin de Borde-Rouge - Auzeville BP27, 31326 Castanet-Tolosan          LOCUS        BX511434              528 bp   mRNA    linear   EST 25-JUN-2003
             cedex, FRANCE                                                            DEFINITION   BX511434 Soares mouse p3NMF19.5 Mus musculus cDNA clone
             Tel: 33 (0) 5.61.28.51.14                                                             IMAGp998J06708 ; IMAGE:314213, mRNA sequence.
             Fax: 33 (0) 5.61.28.53.08                                                ACCESSION    BX511434
             Email: tosser@toulouse.inra.fr                                          VERSION      BX511434.1   GI:32242640
             Clone distribution: AGENAE Resource centre, Francois PIUMI,             KEYWORDS     EST.
             Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et'Etude du          SOURCE       Mus musculus (house mouse)
             genome (LREG), Domaine de Vilvert, 78352, Jouy-en-Josas cedex,           ORGANISM     Mus musculus
             FRANCE, +33 (0) 1.34.65.28.02, +33 (0) 1.34.65.22.73                                 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Sequence cleaned of vector, adaptator and repetitions. Contact us                    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
             at sigenasupport@jouy.inra.fr to obtain the chromatogram of this         REFERENCE    1 (bases 1 to 528)
             sequence.                                                                AUTHORS      Heil,O., Ebert,L., Neubert,P., Peters,M., Radelof,U., Schneider,D.
             Plate: 0035 row: h column: 22.                                                       and Korn,B.
FEATURES     Location/Qualifiers                                                      TITLE        Mouse UnigeneSet - RZPD2
source       1..630                                                                   JOURNAL      Unpublished (2003)
             /organism="Sus scrofa"                                                   COMMENT      Contact: Ina Rolfs
             /mol_type="mRNA"                                                                      RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
             /db_xref="taxon:9823"                                                                 Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
             /clone="scac0035i.h.22"                                                               RZPD; IMAGp998J06708.
             /tissue type="mixed"                                                                  RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
             /clone_lib="Sus Scrofa library (scac)"                                                Mouse Unigeneset - RZPD2 (RZPDLIB No.981)
             /note="vector: pTT3D-pac vector; tissues: adipose tissue,                             http://www.rzpd.de/cloneCards/cgi-
             brain, kidney, liver, muscle, ovary, testis, heart,                                  bin/showlib.pl.cgi/response?libNo=981 Contact: Ina Rolfs
             hypothalamus, pancreas, skin, spleen, thymus, placenta,                              RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
             pituitary gland, seminal vesicle, small intestine,                                   Heubnerweg 6, D-14059 Berlin, Germany
             uterus, adrenals, bulbo uretral gland, cerebral trunk,                               Tel: +49 30 32639 101
             epididymis, female gonad, gall-bladder, hippocampus,                                Fax: +49 30 32639 111
             large intestine, male gonad, melanocytes, stomach, udder"                           www.rzpd.de
ORIGIN                                                                                            This clone is available royalty-free from RZPD;
                                                                                                  contact RZPD (clone@rzpd.de) for further information. Seq primer:
Query Match    35.0%;   Score 401; DB 13;   Length 630;                                           T7, Primer sequence: TAATACGACTCACTATAGG.
. Best Local Similarity  82.2%;  Pred. No. 2.5e-91;                                   FEATURES     Location/Qualifiers
Matches 488; Conservative   0; Mismatches 96; Indels 10; Gaps 2;                     source       1..528
                                                                                                  /organism="Mus musculus"
Qy    1 ATGGCTCCGGCGAAAAGGAAAGGGAGGCGGGGAGGGCCCAGCCAAGAGCGCCCTCC-CGAAGAT 59                        /mol_type="mRNA"
        ||||||||||| ||| ||||||||||||||| ||||| |||||||||||||||||||||                               /db_xref="taxon:10090"
Db    37 ATGGCCCCGGCGAGAAAGGGGAAGAGGGGAGGGAGGGGGAAGAGAGTGCCCTCCNCGGAAGGT 96                         /clone="IMAGp998J06708 ; IMAGE:314213"
                                                                                                  /dev_stage="19.5 dpc total fetus"
Qy   60 ACGCACAGCCACCCTGGTCATCAGTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAA 119                              /lab_host="DH10B (ampicillin resistant)"
        |||||| ||||||||||||||||| ||||||| ||||||||||| ||||||||||||||                               /clone_lib="Soares mouse p3NMF19.5"
Db   97 ACGCACAGCCACCCTGGTTATCAACTTGGCCCGGGGCTGCGCAGCAGTGGGCAAATGAGAA 156

Qy  120 CAGCATCAGGCAGGCCCAGGAGCCCACAGCGCTGCTGCCGCGGGAGGGGACCCAGGACTCACC 179
        |||| ||||||||| |||||| ||| ||||||| ||||||||
```

```
/note="Vector: pT7T3D (Pharmacia) with a modified
polylinker; Site 1: Not I, Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCGCCATTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
```

ORIGIN

```
Query Match          34.3%;  Score 392.6;  DB 13;  Length 528;
Best Local Similarity 84.1%;  Pred. No. 3.1e-89;
Matches  443;  Conservative   0;  Mismatches   84;  Indels   0;  Gaps   0;

Qy  406 GGTGTGCTTGAACCTGGGCAGGCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGC 465
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    2 GGCGTGTCTGAAGCTATTCAGCCAGAGAATGACATTGACAGAATCCTTCTTAGTCACGAC 61

Qy  466 TCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGA 525
            |||| |||||||||||||||||||||| |||||| ||||||||||||||||||||||||
Db   62 TCGCCAACACGGAGGAAGAAAATGCACCAACCTGGTGTCTGACCTGACCAAAGGCTCGAAA 121

Qy  526 GTGATGGAGCAGGAGGAGCCCACATGAGGAGTGACAGCGTAGACACAGAGGACAGCGGGC 585
            ||||||||||||||||||||| ||| ||||||||||||||||||||||||||||||||
Db  122 GTGATGGAACAGGAAGAGCCCCACGTGGAAGAGTGACACGGTAGACACAGAGGACAGTGGC 181

Qy  586 TATGGAGGAGGAGGCTGAGGAGGGCCCGAGCAGCAGGAGTGCAGGTGGGCGTGTGGTCAGG 645
            ||||| |||||| ||||||||||| |||||||||||||||||| |||||||||||||||
Db  182 TACGGAGGGGGATATGGAGGAGGGCCTGAGCAAGATGCCAGCGCCCTGTTGGCTCCTGCCAGG 241

Qy  646 ATCAAGCGCCCCTTGCCCTCCCAGTAAACAGATTTACAGAGAAAACTCAACTGCAAAGCC 705
            |||| ||||||||||||||||||| ||||||||||||| |||||| ||||||||||||||
Db  242 ATCAAACGCCCCTTGCTCTCCCAGGCAAACAGGTACTCTGAGACACTCAACTGTAAGGCC 301

Qy  706 CAACAGAAATATAGCCCAGTGGCAACTTGAAAGGAGATGGCAGCAGTGGGCTGATGAA 765
            ||||||||||||| ||||||||||||||||||||||||||||||| |||||||||||||
Db  302 CATCGGAAATACAGCCAAGTGACAACTTGAAAGGGGAGTGGCAGCAGTGGGCCGATGAA 361

Qy  766 CACATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTTACGAGCTGGCCATG 825
            ||||| |||||||||||||||||||||||||||| ||||||||| |||||||||||||||
Db  362 CAGTCCAGTCCCAGAAGCTCAATCCCTTCAGTGACGAATTTGACTATGATCCTAGCCATG 421

Qy  826 TCCACCCGCCTACCACAAAGGAGGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACT 885
            ||||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db  422 TCCACTCGGCTCCACAAGGAGGACGAGGGCTATGGCCGCCCCAAAGAAGGGAAGCAAGACA 481

Qy  886 GCTGAAAGGGCCAAGCGTGCTGAGGAGCACATCCACAGGGAAATGAT 932
            |||||||||||||||||||||||||||| |||||| |||||||||||
Db  482 GCTGAAAGGGCCAAGCGAGCGGAAGAGCACCATCTATCGGGAAATTAT 528
```

Search completed: May 5, 2004, 03:03:18
Job time : 2878 secs

```
           GenCore version 5.1.6
      Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:   May  4, 2004, 08:51:39 ; Search time 60.5 Seconds
                                   (without alignments)
                                   10501.133 Million cell updates/sec

Title:          US-10-644-659A-1
Perfect score:  2058
Sequence:       1 atgctccgggcgaaaaga.........tgattacgctaccaagtga 1146

Scoring table:  BLOSUM62
      Xgapop 10.0 ,  Xgapext 0.5
      Ygapop 10.0 ,  Ygapext 0.5
      Fgapop  6.0 ,  Fgapext 7.0
      Delop   6.0 ,  Delext  7.0

Searched:  1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters:    2276240

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10644659/runat_04052004_084903_28592/app_query.fasta_1.1287
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10644659 @CGN_1_1_13_@runat_04052004_084903_28592
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
   1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
   2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
   3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
   4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
   5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
   6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
   7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
   8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
   9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
  10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
  11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
  12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
  13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
  14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
  15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
  16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
  17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
  18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	%	Length	DB	ID	Description
1	2029	98.6	381	15	US-10-104-047-3169	Sequence 3169, Ap
2	647	31.4	121	9	US-09-864-761-42191	Sequence 42191, A
3	190	9.2	37	9	US-09-864-761-42192	Sequence 42192, A
4	142	6.9	19695	15	US-10-084-846A-3	Sequence 3, Appli
5	134.5	6.5	19662	15	US-10-084-846A-6	Sequence 6, Appli
6	133	6.5	19608	15	US-10-084-846A-8	Sequence 8, Appli
7	130.5	6.3	1739	15	US-10-341-434-61	Sequence 61, Appl
8	129.5	6.3	19723	15	US-10-084-846A-5	Sequence 5, Appli
9	126.5	6.1	93	12	US-10-425-114-63655	Sequence 63655, A
10	126.5	6.1	19652	15	US-10-084-846A-7	Sequence 7, Appli
11	125.5	6.1	645	12	US-10-425-114-45380	Sequence 45380, A
12	124.5	6.0	339	12	US-10-425-114-45434	Sequence 45434, A
13	123.5	6.0	422	12	US-10-425-114-70922	Sequence 70922, A
14	123.5	6.0	460	12	US-10-425-114-71620	Sequence 71620, A
15	122	5.9	293	14	US-10-029-386-33391	Sequence 33391, A
16	121.5	5.9	506	14	US-10-156-761-12648	Sequence 12648, A
17	121	5.9	284	14	US-10-029-386-33066	Sequence 33066, A
18	121	5.9	488	12	US-10-365-742-20	Sequence 20, Appl
19	121	5.9	1182	14	US-10-024-368-6	Sequence 6, Appli
20	120.5	5.9	1019	12	US-10-276-774-2298	Sequence 2298, Ap
21	120	5.9	2803	15	US-10-415-187-5	Sequence 5, Appli
22	120	5.8	485	14	US-10-091-752A-14	Sequence 14, Appl
23	120	5.8	485	14	US-10-136-547-14	Sequence 14, Appl
24	119.5	5.8	375	14	US-10-155-761-14622	Sequence 14622, A
25	119.5	5.8	442	12	US-10-425-114-40452	Sequence 40452, A
c 26	119	5.7	19695	15	US-10-084-846A-3	Sequence 3, Appli
27	118.5	5.8	311	15	US-10-425-114-61343	Sequence 61343, A
c 28	117.5	5.6	524	15	US-10-289-762-52	Sequence 52, Appl
29	116.5	5.7	91	12	US-10-422-599-237720	Sequence 237720,
30	116.5	5.7	336	12	US-10-425-114-56601	Sequence 56601, A
31	116.5	5.7	685	15	US-10-104-047-2916	Sequence 2916, Ap
32	115.5	5.6	656	15	US-10-108-260A-2611	Sequence 2611, Ap
33	115.5	5.6	19725	15	US-10-084-846A-4	Sequence 4, Appli
34	115	5.6	115	15	US-10-384-919-2	Sequence 2, Appli
35	114.5	5.6	300	12	US-10-425-114-64837	Sequence 64837, A
36	114.5	5.6	336	12	US-10-425-114-68820	Sequence 68820, A
37	114	5.5	383	12	US-10-425-114-53657	Sequence 53657, A
38	113.5	5.5	324	12	US-10-425-114-53603	Sequence 53603, A
39	113.5	5.5	361	12	US-10-425-114-59602	Sequence 59602, A
40	113.5	5.5	362	12	US-10-425-114-58177	Sequence 68177, A
41	113.5	5.5	367	12	US-10-425-114-65068	Sequence 65068, A
42	113	5.5	324	12	US-10-425-114-47752	Sequence 47752, A
43	113	5.5	1593	11	US-09-981-151A-30	Sequence 30, Appl
44	113	5.5	1593	15	US-10-295-027-1317	Sequence 1317, Ap
45	113	5.5	2038	16	US-10-433-794-18	Sequence 18, Appl

ALIGNMENTS

```
RESULT 1
US-10-104-047-3169

; Sequence 3169, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
;  APPLICANT: HELIX RESEARCH INSTITUTE
;  TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
;  FILE REFERENCE: H1-A0105
;  CURRENT APPLICATION NUMBER: US/10/104,047
;  CURRENT FILING DATE: 2002-03-25
;  PRIOR APPLICATION NUMBER:
;  PRIOR FILING DATE:
;  NUMBER OF SEQ ID NOS: 4096
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3169
;  LENGTH: 381
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-104-047-3169

Alignment Scores:
Pred. No.:           1.11e-158          Length:        381
Score:               2029.00            Matches:       381
Percent Similarity:  100.00%            Conservative:  0
```

```
Best Local Similarity:   100.00%      Mismatches:  0
Query Match:              98.59%       Indels:      0
DB:                                    Gaps:        0

US-10-644-659A-1 (1-1146) x US-10-104-047-3169 (1-381)

Qy     1 ATGGCTCCGGCGAAAAGAGAAAAGCGGGAGGGGCCAGCCAAGAGGCGCCTCCGGAAGATA  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle  20

Qy    61 CGGACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCCAGCAGTGGGCGAATGAGAAC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn  40

Qy   121 AGCATCAGGCAGGCAGGCCTACAGGCTGGCGTGCCCGGAGGACCCAGGACTCACCT  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyThrGlnAspSerPro  60

Qy   181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACCACGAAAGCTCAGAGTGCCCCAAAG  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GlnAlaProLysProIleThrProProThrSerHisGlnLysAlaGlnSerAlaProLys  80

Qy   241 TCGCCACCCCGCCTGCCAGAGACGACATGGACAAAGTCGACAGAGAAAGCCCCTGAG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 SerProProArgLeuProGluGlyHisGlyAspGlyGlnSerSerGluLysAlaProGlu  100

Qy   301 GTTTCTCACATCAAAAAGAAGAGAGTGTCCAAAACGGTGCTCAGCAAGACTTACGAGAGA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ValSerHisIleLysLysLysArgValSerLysThrValLeuSerLysThrTyrGluArg  120

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACCAGTACGAGAGGATGCGGTGTGCTGAACCT  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GlyGlyAspValSerHisLeuSerHisGlnTyrGluArgGlyAspGlyValLeuGluPro  140

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGCGGGAGG  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 GlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisGlySerProThrArgArg  160

Qy   481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGGCTGGAGAGTGATGGGAGCAGGAG  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyLeuGluSerGlyValMetGluGlnGluGlu  180

Qy   541 GAGCCCACATGGAGGAGTGACAGCGGACACAGAGGACGGCTATGGAGGAGAGGCT  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GluProThrTrpArgSerAspSerValAspSerGlyTyrGlyGlyGluLeuAla  200

Qy   601 GAGGAGAGGCCCGAGCCAGGAGGAGTGCTGGTCTGAGGATCAAGCGCCCCTTG  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GluGluArgProGluGluProGlnAspGlyValGlnValAlaValAlaArgProLeu  220

Qy   661 CCCTCCCAGTGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ProSerGlnValAsnArgPheThrGluLysLeuAsnCysLysSerAlaGlnGlnLysTyrSer  240

Qy   721 CCAGTGGCCAACTTGAAAGGAGGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAG  780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ProValGlyAsnLeuLysGlyGlyTrpGlnGlnTrpAlaAspGluHisIleGlnSerGln  260

Qy   781 AAGCTCAATCCTTCAGTGAAGGAGTTTGATTACGAGCCCATGTCCACCCGCCTACAC  840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 LysLeuAsnProPheSerGluGluPheAspTyrGluAlaMetSerThrArgLeuHis  280

Qy   841 AAAGGAGAATGAGGGCTATGGCGCCCCAAAGAAGAACCAAAACTGCTGAAAGGCCAAG  900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 LysGlyGluGluGlyTyrGlyArgProLysGlyThrLysThrAlaGluArgAlaLys  300

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGACAATG  960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ArgAlaGluGluHisIleTyrArgGluMetMetAspMetCysPheIleIleCysThrMet  320

Qy   961 GCTCGCCACAGACGAGATGGCAAGATCCAGTTACTTTGGAGATCTCTTGACAGATAC  1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 AlaArgHisArgArgAspProGlyLysIleGlnValThrPheGlyAspLeuPheAspArgTyr  340

Qy  1021 GTTCGTATTCAGATAAGTAGTGGGCATTCTCATGCTGCCAGGAAAACATGGACTGGTA  1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341 ValArgIleSerAspLysSerValValGlyValIleLeuMetLeuArgAlaArgLysHisGlyLeuVal  360

Qy  1081 GACTTTGAAGAGGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AspPheGluGluGluMetLeuTrpGlnGlyArgAspAspHisValValIleThrLeuLeu  380

Qy  1141 AAG 1143
          |||
Db   381 Lys 381
```

```
RESULT 2
US-09-864-761-42191
; Sequence 42191, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F(
;   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aeomica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   CURRENT FILING DATE:   2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE:   2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE:   2000-05-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE:   2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE:   2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE:   2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00667
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00664
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00669
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00665
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00668
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00663
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE:   2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE:   2001-01-29
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE:   2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE:   2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE:   2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42191
;   LENGTH: 121
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC023344.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
```

```
,   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
,   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
,   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
,   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
,   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
,   OTHER INFORMATION: EST_HUMAN HIT: BF083249.1, EVALUE 8.50e-01
,   OTHER INFORMATION: SWISSPROT HIT: P03189, EVALUE 8.80e+00
US-09-864-761-42191

Alignment Scores:
Pred. No.:
Score:                     8.54e-45          Length:            121
Percent Similarity:        647.00            Matches:           121
Best Local Similarity:     100.00%           Conservative:        0
Query Match:               100.00%           Mismatches:          0
DB:                        31.44%            Indels:              0
                           9                 Gaps:                0

US-10-644-659A-1 (1-1146) x US-09-864-761-42191 (1-121)

Qy 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTCTCACATCAAAAGAAA 321
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GlyHisHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys 20

Qy 322 GAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGGGACGTGAGCCACCTC 381
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyArgGluValSerHisLeu 40

Qy 382 AGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGCAGCCAGAGAATGACATT 441
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy 442 GACAGAATCCTCCCACAGCCACGGCTCCCCACGCGCGAGGAGAAAATGTGCCAACCTGGTG 501
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgLysCysAlaAsnLeuVal 80

Qy 502 TCTGAGCTAACCAAGGGCTGGAGTCATGAGCAGCAGGAGCCCCATGGAGGAGGAGTGAC 561
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGluProThrTrpArgSerArgAsp 100

Qy 562 AGCGTAGACACAGAGGACAGCAGCCGCTATGGAGGAGGAGGCTGAGGAGGAGGCCCGAGCAGGAT 621
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 SerValAspThrGluAspSerSerGlyTyrGlyGlyGlyGlyAlaGluGluLeuArgArgArgProGluGluGlnAspAsp 120

Qy 622 GGA 624
       |||
Db 121 Gly 121

RESULT 3
US-09-864-761-42192
; Sequence 42192, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
   APPLICANT: Penn, Sharron G.
   APPLICANT: Rank, David R.
   APPLICANT: Hanzel, David K.
   APPLICANT: Chen, Wensheng
   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
   FILE REFERENCE: Aeomica-X-1
   CURRENT APPLICATION NUMBER: US/09/864,761
   CURRENT FILING DATE: 2001-05-23
   PRIOR APPLICATION NUMBER: US 60/180,312
   PRIOR FILING DATE: 2000-02-04
   PRIOR APPLICATION NUMBER: US 60/207,456
   PRIOR FILING DATE: 2000-05-26
   PRIOR APPLICATION NUMBER: US 09/632,366
   PRIOR FILING DATE: 2000-08-03
   PRIOR APPLICATION NUMBER: GB 24263.6
   PRIOR FILING DATE: 2000-10-04
   PRIOR APPLICATION NUMBER: US 60/236,359
   PRIOR FILING DATE: 2000-09-27
   PRIOR APPLICATION NUMBER: PCT/US01/00666
   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00667
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00664
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00669
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00665
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00668
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00663
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00662
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00661
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: PCT/US01/00670
,   PRIOR FILING DATE: 2001-01-30
,   PRIOR APPLICATION NUMBER: US 60/234,687
,   PRIOR FILING DATE: 2000-09-21
,   PRIOR APPLICATION NUMBER: US 09/608,408
,   PRIOR FILING DATE: 2000-06-30
,   PRIOR APPLICATION NUMBER: US 09/774,203
,   PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 42192
   LENGTH: 37
   TYPE: PRT
   ORGANISM: Homo sapiens
   FEATURE:
   OTHER INFORMATION: MAP TO AC023344.2
   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
   OTHER INFORMATION: EST_HUMAN HIT: AV696020.1, EVALUE 5.60e+00
US-09-864-761-42192

Alignment Scores:
Pred. No.:                 3.23e-07          Length:             37
Score:                     190.00            Matches:            36
Percent Similarity:        97.30%            Conservative:        0
Best Local Similarity:     97.30%            Mismatches:          1
Query Match:               9.23%             Indels:              0
DB:                        9                 Gaps:                0

US-10-644-659A-1 (1-1146) x US-09-864-761-42192 (1-37)

Qy 260 AAGGACAATGGAGGATGGACAGACAAAGCTCAGAGAGAAGCCCCTGAGGTTCTCACATCAAAAGA 319
        |||||||||||||||||||||||||||||||||||||||||||||
Db   1 LysAspMetGluMetAspLysAlaGlnArgLysProLeuArgPheLeuThrSerLysArg 20

Qy 320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGAGGGGGACG 370
        ||||||||||||||||||||||||||||||||||||||||||
Db  21 LysArgCysProLysArgTrpSerArgAlaArgLeuMetArgGluGluGlyThr 37

RESULT 4
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US2004006026A1
GENERAL INFORMATION:
   APPLICANT: WEITNAUER, GABRIELE
   APPLICANT: MUHLENWEG, AGNES
   APPLICANT: TREFZER, AXEL
   APPLICANT: BECHTHOLD, ANDREAS
   TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
   FILE REFERENCE: 1974-005
   CURRENT APPLICATION NUMBER: US/10/084,846A
   CURRENT FILING DATE: 2003-02-25
```

```
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR APPLICATION FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR APPLICATION FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
;  SEQ ID NO 3
;  LENGTH: 19695
;  TYPE: PRT
;  ORGANISM: Streptomyces viridochromogenes
;  FEATURE:
;  OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
;  OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3

Alignment Scores:
Pred. No.:               0.0127          Length:          19695
Score:                   142.00          Matches:         97
Percent Similarity:      32.56%          Conservative:    30
Best Local Similarity:   24.87%          Mismatches:      161
Query Match:             6.90%           Indels:          104
DB:                      15              Gaps:            17

US-10-644-659A-1 (1-1146) x US-10-084-846A-3 (1-19695)

Qy      2  TGGCTCCGGGCGAAAAGGAAAGCGGGGAGGGCCCAGCCAGAGGGCCCTCCGGAAGATAC  61
              |||  |||||||||                                  ::
Db   8295  TrpArgAlaAlaLysArgThrCysAla-----------------AlaProSerSerSerTrp  8309

Qy     62  --------GCAGCAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGA  112
                   :||||||  |||||   |||||   |||  ||:|||::|||||
Db   8310  ArgCysThrAlaSerProThrTrpAsnSerGlySerProThr--TrpSerArgTrpSerS  8329

Qy    113  ATGAGAACAGCATC----------AGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCGG  160
              ::                     ||||||
Db   8329  erProGlySerAlaThrSerThrProArgSerAlaGlnGlySerSerThrThrCysSerG  8349

Qy    161  GAGGGACCCAGGAGACTCACCT--CAAGTCCTAAACCAATC------------------  198
              |||||||||||    ||||||        ||
Db   8349  lyProGlySerCysSerProThrSerAlaProSerProThrAlaSerThrSerProProT  8369

Qy    199  ----------ACACCCCTACTTCACACCAGAAAGTCAGAGTGCCCCAAAGT  241
                        |||||          :::
Db   8369  hrThrGlyAlaAlaArgThrGluProThrSerThrValArgGlyProProArgA  8389

Qy    242  CGCCACCCCGCCTGCCAGAGGACATGGAGGACAAAGCTCAGAGAAAAGCCCCTGAGG  301
              ||||||||||||       :::
Db   8389  rgProProArgAlaProGlyProAlaProGlnAsnArgArgArgAlaLeuAlaGlyProAlaV  8409

Qy    302  TTTCTCCA----------CATCAAAAAGAAAGAGAGTGTCCAAAACG  336
              |::||||                :::
Db   8409  alAlaAspAlaArgProGlyAlaGlyGlyLeuHisArgCysArgGlyAlaAlaArgAlaG  8429

Qy    337  GTGGTCAGCAAGACTTACGAGAGGAGGAGGGAGCGTGAGCCACCTCCAGCCACA-GGTACCA  395
              ||  ||||||  :::                               |||
Db   8429  lyHisGlyArgGlyLeuArgHisArgArgGly----------LeuProGlnGlyArgA  8445

Qy    396  GAGGGATGCTGGTCGTGCCTGAACCTGGCCAGCCCAGAGAATGACATTGACAGAATCCTCCA  455
              ||:||||||||||                            :::|| |||
Db   8445  rgAlaProLeuValThrLeuProGlyArgGlyGluArgArgThrHisProProArgArgH  8465

Qy    456  CAGCCACGGCTCCCCAACCGGGA----------GGAGAAAAGTGCCAACCTGGTGTC  503
              |||||||                              |||
Db   8465  isTrpThrGlyProArgSerGlySerAspThrAlaSerArgThrArgAlaA  8485

Qy    504  TGAGCTAACCAAGGGCTGGAGGCTGGAGTGATGGAGCAGCCCCATGGAGGAGTGACAG  563
              ||||                     |||||||:::|||
Db   8485  rgPheSerProArgThrArgSerProValSerArgArgAlaProAsnGlySerSerProA  8505

Qy    564  CGTAGACACAGAGG----------ACAGCGGCTATGGAGGAGAGGCTGAGGA  605
              ||||||                    |||||||:::|||
Db   8505  laSerProThrArgAlaCysAlaProThrThrArgGlyThrThrSerProSerArgAlaMetS  8525

Qy    606  GAGGCCCGAGCAGGAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCCTC  665
              |||                   |||     :::                       |||
Db   8525  erValProSerThrAlaCysThr---ThrTrpArgSerArgThrGlyPro--------  8541

Qy    666  CCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACGAGAAATATAGCCCAGT  725
              ::|                                                |||
Db   8542  ----------ArgAsnGlyThrCysThrAlaArgAsn-----------T  8552

Qy    726  GGGCAACTTGAAAGGGAGGATGGCAGCAGTGGGCGTGATGAACACATACAATCCCAGAAGCT  785
              ||                                       :::          ||||||||
Db   8552  rp----------ArgSerProProArgSer-  8559

Qy    786  CAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACACAAAGG  845
                                                  TrpThrAlaArgSerCysSerProProThrProArgSerS  8572
Db   8560  ----------

Qy    846  AGATGAGGGCTATGGCCGCCCCAAAGAGGAGGACCAAAACTGCTGAAAGGGCCAAGCGTGC  905
              |:::::::                               |||                :::
Db   8572  erSerArgSerArgThrGlyProArgArgCysHisGlyThrValThrAlaSerThrSerA  8592

Qy    906  TGAGGGAGCACATCT----------ACAGGGAAATGAT  932
              |:::::                     |||
Db   8592  laArgAlaArgSerAlaIleAlaSerAlaTrpThrThrProAspArgThrThrAspAlaT  8612

Qy    933  GG----------ACATGTGCTTCATTATCTGCACAATGG----CTCG  965
              ||                        :::              ::::
Db   8612  rpArgArgPheArgAspArgThrCysSerProThrThrProArgTrpArgArgCysAlaA  8632

Qy    966  CCACAGACGAGATGGCAAGATCCAGG  991
              |||
Db   8632  rgProAspProTrpPheArgSerArg  8640

RESULT 5
US-10-084-846A-6
;  Sequence 6, Application US/10084846A
;  Publication No. US2004006026A1
;  GENERAL INFORMATION:
;  APPLICANT: WEITNAUER, GABRIELE
;  APPLICANT: MUHLENWEG, AGNES
;  APPLICANT: TREFZER, AXEL
;  APPLICANT: BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
SEQ ID NO 6
;  LENGTH: 19662
;  TYPE: PRT
;  ORGANISM: Streptomyces viridochromogenes
;  FEATURE:
;  OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.
;  OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6

Alignment Scores:
Pred. No.:               0.0528          Length:          19662
Score:                   134.50          Matches:         91
Percent Similarity:      34.26%          Conservative:    20
Best Local Similarity:   28.09%          Mismatches:      106
Query Match:             6.54%           Indels:          107
DB:                      15              Gaps:            13

US-10-644-659A-1 (1-1146) x US-10-084-846A-6 (1-19662)

Qy      8  CGGGCGAAAAGGAAAGCGGGGAGGGCCCAGCCAGAGGGCCCTCCGGAAGATACGCACAG  67
              ||||||::::||                  ::|||||           :::|||||
Db  12856  ArgAlaArgArgGlySerGlyArgArgProArg---ProSerGlyArgProAlaAla  12874
```

```
;  APPLICANT: BECHTHOLD, ANDREAS
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE: 2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
;  SEQ ID NO 8
;  LENGTH: 19608
;  TYPE: PRT
;  ORGANISM: Streptomyces viridochromogenes
;  FEATURE:
;  OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
;  OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-8

Alignment Scores:
Pred. No.:                              Length:               19608
Score:                  0.0702          Matches:              95
Percent Similarity:     133.00          Conservative:         27
Best Local Similarity:  28.99%          Mismatches:           153
Query Match:            22.57%          Indels:               146
DB:                     6.46%           Gaps:                 19
                        15

US-10-644-659a-1 (1-1146) x US-10-084-846A-8 (1-19608)
```

DB: 15   Gaps: 10

US-10-644-659A-1 (1-1146) x US-10-341-434-61 (1-1739)

```
Qy    109  GCCAATGAGAACAGCATCAGGCAGGCCCAGAGCTTCACAGGCTGGCTGCCGGAGGGACC  168
                ::::|||       :::|||  :::|||  ||||||      :::|||
Db     17  SerAsnAlaSerSerHisSerAlaSerGluGluAlaSerGlySerAspSerGlySerGln   36

Qy    169  CAGGACTTCACCTCAAGCTCTAAACCAATCACACCCCTACTTCACAC---CAGAAAGCT  225
                |||  :::|||                                  ||| :::|||
Db     37  SerGluSerGlyGlnGlySer---------AspProGlySerGlyHisGlySerGluSer   53

Qy    226  CAGAGTGCCCAAAGTGCCACCCGCTGCCAGAAGGACATGGAGATGGACAAAGTCA  285
                |||||    :::|||  :::|||
Db     54  AsnSerSerSerGluSerGlnSerGlnSerGluSerGluSerAlaGlySer   73

Qy    286  GAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAGAGGTGTCAAAACGTGGTCAGC  345
                |||||||  ||||||
Db     74  LysSerGlnProValLeuProGluAlaValAlaGlyLysProAlaSerLys-------   90

Qy    346  AAGACTTACGAGAGGAGGGACGTGAACATGGGCCACCTCAGCCACCAGTACGAGAGGGATGCT  405
                ||||||       :::
Db     91  -------GluArgIleAlaAspValLysLysMetTrpGluGluTyrProAspValTyr  107

Qy    406  GGTGTGCTTGAACCTGGGCAGCCAGGAGAATGACATTGACAGAATCCTC----------  453
                |||||       :::      :::
Db    108  GlyValArgArgAsnArgSerArgGlnGluProSerArgPheAsnIleLysGluGlu  127

Qy    454  ----------CACCGCCACGGCTCCCAACCGGGAGGAGAGAAAAATGTCCACCTGCTG  501
                      ||||||||
Db    128  AlaSerSerGlySerGluSerGlySerProLysGluArg----------------  140

Qy    502  TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAGGAGCCCACATGAGGAGTGAC  561
                |||||||                   :::
Db    141  ----------GlyGlnArgGlnLeuLysLysGlnGluLysTrpLysGlnGluPro  155

Qy    562  AGGGTAGAACACAGGACAGCAGCCGGCTATGGAGGAGGAGGCTGAGGAGGCCCGAGCAGGAT  621
                ||||||       ||||||||
Db    156  SerGluAspGluGlnGluGlnGlyThrSerAlaGluGlu----------ProGluGln--  172

Qy    622  GGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCTCCCCAGGTAAACAGAATT  681
                ||||||||  :::|||
Db    173  ----------LysLysValLysAlaAArgArgProValPro---------ArgArgThr  185

Qy    682  ACACAGAAAACTCAACTGCAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGG  741
                |||           |||       |||
Db    186  ValProLysProArgValLysLysGlnProLys---------ThrGlnArgGlyLysArg  202

Qy    742  AGATGGCAGCAGGCTGGGCTGATGAA-----------------------------  765
                ||||||
Db    203  LysLysGlnLysAspSerSerAspGluAspAspAspAspGluAlaProLysArgGlnThr  222

Qy    766  CACATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATG  825
                :::::::       :::
Db    223  ArgArgArgAlaAlaLysAsnValSerTyrLysGluAspAspAspPheGluThrAspSer  242

Qy    826  TCCACCCGCTACACAAAGGAGGATGAGGGCTATGGCGCGCCCCCAAAGAAGAGGAACCAAAACT  885
                :::
Db    243  AspAspLeuIleGluMetThrGlyGluGlyValAspGluGlnGlnAspAsnSerGluThr  262

Qy    886  GCTGAAAGG  894
           |||
Db    263  IleGluLys  265
```

RESULT 8
US-10-084-846A-5
; Sequence 5, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS

```
Qy    500  TGTCTGAGCTAACCAAGG-------------------------------------  517
           |||||
Db   5380  ArgArgCysTrpProArgTrpSerArgAsnProArgProCysAlaThrSerTrpArgThr  5399

Qy    518  -------GCTGGAGAGTGATGGAGCAGGAGGAGGCCCACATGAGGAGTGACAGCG  565
                |||||||       ||||||
Db   5400  AlaSerThrSerArgAlaTyrGlyTrpSerArgThrSerPro---------ThrSer  5415

Qy    566  TAGACACAGAGACAGGACGCGGCTATGGAGGCTGAGGAGAGCCCG---AGCAGGATG  622
           |||       |||
Db   5416  ProThrAlaSerThrAlaProAlaThrThrSerProAlaThrProAlaSerSerThr  5435

Qy    623  GAGTGC--------AGGTGGCTGTGGTCAGGATCAAGCGCC--------CCT  658
           |||:::               :::|||:::|
Db   5436  ArgCysCysProAlaSerThrTrpArgSerThrArgAlaAlaCysTrpProArgArgPro  5455

Qy    659  TGCCCTCCAGGTAAACAGATTACAGAGA----------------AACTCA  694
                |||  |||||||
Db   5456  CysSerSerProArgThrArgArgThrArgArgArgProGlyArgSerThrArgSerSer  5475

Qy    695  ACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGGCAGCAGT  754
           ||||||       |||                 :::  |||
Db   5476  ThrAlaThrProThrSerGlyCysSerProTrpCysPro----------------  5488

Qy    755  GGGTGATGAACAACATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACG  814
                      :::      :::                :::
Db   5489  --------------AlaCysThrSerAsnArgProAlaArgProThrProThrSer  5501

Qy    815  AGCTGGCCATGTCCACCCGCTACACAAAGGAGGATGAGGGCTATGGGCCGCCCCAAAGAAG  874
           ||||||||                                       ::|||||||||:::
Db   5502  SerTrpProThrArgPro------------------ArgSerAlaProArg--  5512

Qy    875  GAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGG  934
                                                      ||||||||||:::
Db   5513  ----------------------ProSerThrSerArgSerThrAlaArg-------Trp  5522

Qy    935  ACATGTGCTTCATTATCTGCACAATGG-----CTCGCCACAGACGAGATGGCAAGATCC  988
                |||      :::            |||        |||||||:::
Db   5523  ProSerProAsnTrpSerArgAspTrpProThrSerThrProSerProAlaArgThr  5542

Qy    989  AGG  991
           |||
Db   5543  Arg  5543
```

RESULT 7
US-10-341-434-61
; Sequence 61, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR APPLICATION FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-61

Alignment Scores:
Pred. No.:   0.0642   Length: 1739
Score: 130.50%   Matches: 63
Percent Similarity: 37.10%   Conservative: 42
Best Local Similarity: 22.26%   Mismatches: 123
Query Match: 6.34%   Indels: 55

```
;  TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
;  FILE REFERENCE: 1974-005
;  CURRENT APPLICATION NUMBER: US/10/084,846A
;  CURRENT FILING DATE:      2003-02-25
;  PRIOR APPLICATION NUMBER: PCT/EP01/09815
;  PRIOR FILING DATE: 2001-08-24
;  PRIOR APPLICATION NUMBER: DE 101 09 166.4
;  PRIOR FILING DATE: 2001-02-25
;  NUMBER OF SEQ ID NOS: 120
;  SOFTWARE: PatentIn Ver. 3.2
;  SEQ ID NO 5
;  LENGTH: 19723
;  TYPE: PRT
;  ORGANISM: Streptomyces viridochromogenes
;  FEATURE:
;  OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.
;  OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.
US-10-084-846A-5

Alignment Scores:
Pred. No.:                0.137        Length:        19723
Score:                  129.50         Matches:          99
Percent Similarity:      28.82%        Conservative:     33
Best Local Similarity:   21.62%        Mismatches:      147
Query Match:              6.29%        Indels:          179
DB:                                    Gaps:             20

US-10-644-659A-1 (1-1146) x US-10-084-846A-5  (1-19723)

Qy     12  CGAAAAGGAAAGCGGGGAGGGCCCAGCGACCCTCCGGAGATACGCACAGCCAC   71
               ||| ||::|:|    |||        |||         ||||||
Db     48  ArgArgGlyProArg---ProProGlyHisArgAspProProHisProArgHisGlyHis   66

Qy     72  ----------------------CCTGGTCATCAGCTTGGC-------------------   89
                               |||||
Db     67  ArgGlnArgAlaValHisArgArgArgProGlyProGlnHisGlyAlaGlyProProArg   86

Qy     90  --------CCGAGGTTGGCAGCAGTGGCGGAATGAGAACAGCATCAGGCAGGCCCAGGA  140
                  ||||||                       :::||||||
Db     87  ThrProTrpProArgArgGluAla------ArgLeuGlnGlnGlnGlnGlyGlyArgGly  104

Qy    141  GCCTACAGGCTGCTGCCGGGAGGGACCCAGGACTCACCTCAAGCTCCTAAACCA-----  195
               ||||                   |||:|||   |||:||
Db    105  AlaGluLeuGluGlyAlaLeuGly-----Pro-ProArgProGluValProArgProCysAr  122

Qy    196  --ATCACACCCCTACTTCACACCAGCAAAGCTTCAGACGTCCAAAGTCGGCCA---  246
               ||       |||      :::   :::       ||:||||
Db    122  gAspAlaValLeuProGlyAlaValHisHisArgArgGlyGluProGlnValProGl  142

Qy    247  ---CCCCGCCTGCCAGAAGGACATGGAAGCATGAGAGATGACAAAGCTCAGAGAAAGCCCCTGAGGT  302
               ||                   ||||||
Db    142  nValProAlaProHisArgGlyHisGlnGlnGlnGlyGlyGlyAspArgProGlyProAlaAlaGlu--  161

Qy    303  TTCTCACATCAAAAAAGAAGAGAGGTGTCCAAAACGGTGTCAGCAAGACTTACGAGAGAGG  362
                                                   :::      ||||||
Db    162  -----------------------------------ACGCGGAGGAAAATGTGCCAACCTGGTCTCTGA  168

Qy    363  AGGGGACGTGAGCCACCTGCAGCTACGAGAGGGGATGCTGGTGTGCTTCGAACCTGG  422
               |||||||                   ||||||
Db    168  yGlyAspAlaGlyGlyGlyGluGlyArg------------LeuGlyProAlaThrProGl  184

Qy    423  GCAGCCAGAGAATGACAATTCACAGAATCCTCCACAGCCACGGCTCCCCA-----------  471
                               :::      :::         :::||||||::||
Db    184  ySerArgArgGlnGlnLeuAlaAlaAlaValThrTrpThrGlyGlyProArgProCysGl  204

Qy    472  --------------------ACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGA  506
                          :::||||||            :::
Db    204  yProAlaAlaGlySerSerValArgSerArgArgSerAlaSerLeuArgThrAl  224

Qy    507  GCTAACCAAGGGCTCGAGAGTGATGGAGCAGGAGGAGCCCACATGCGAGGAGGAGGTGACAGCGT  566
               ||||||::||:||                   :::||||||
Db    224  aThrArgProGlyTrpArg------Gly-ArgArgArgAlaProSerGlyValArgArgSerP  242
```

Right column:

```
Qy    567  AGACACAG---AGGACAGCGGCTATGGAGGAGGAGGGCTGGAGGAGAGGC----------CCGA  614
               :::||   ||||                   |||||
Db    242  roArgArgCysArgProProAlaProAlaProAlaProArgArgArgSerValProProProA  262

Qy    615  GCAGGAGGATGCAGGTGGCTGGTGGTCAGGATCAAGCGCCCCTGCCTCCC--------  667
               ||||||                   ||||||
Db    262  laArgSerProGlyArg---------------ProCysAlaProThrProProSerGlyA  277

Qy    668  --------------------------------AGGTAAACAGATTTACAGAGAA  689
                                                :::
Db    277  rgThrArgAlaGlyThrAlaArgCysThrAlaAlaThrThrArgThrAlaAlaSerArgG  297

Qy    690  ACTCAACTGCA------------------------AAGCCCAACAGAAATATAG  719
               |||                   ||||:::
Db    297  lySerArgAlaGlyLyAlaSerArgArgArgArgThrGlyArgArgProAlaArgSerAlaG  317

Qy    720  CCCAGTGGGCAACTTGAAAGGAGGAGATGGCAGC--------------------  751
               |||      |||
Db    317  lyGlyTrpGluThrGlyAlaGlyProArgGlyAlaAlaArgGlyAlaAlaValProP  337

Qy    752  ----------------------------------------------------A  752
               |||
Db    337  roArgAlaThrGlyProHisGlyProTyrGlnGlyValaGlyValaGlyArgAlaGlyTrpS  357

Qy    753  GTGGGCTGATGAACAACATACAAT-----------------  775
               ||||||
Db    357  erGlyArgAlaGlyThrGlyAlaGlyThrArgAlaTrpArgThrGlyArgAlaSerArgArgProGlyP  377

Qy    776  --------------------------------CCAGAAGCTCAATCCTTT  794
                                                 |||
Db    377  roAspArgSerArgArgAlaAArgAsnArgArgArgAlaProHisArgSerArgLeuG  397

Qy    795  CAGTGAAGAGTTGATTACGAGCTGG--------------------  820
               ||:||||      :::   :::::::
Db    397  luValArgSerArgLeuAlaAlaTyrArgArgLeuHisProAlaProHisArgArgProArgA  417

Qy    821  ----------------------------CCATGTCCACCCGCC--------  835
                                              |||||||
Db    417  rgGluArgHisProAlaHisProAlaProAlaThrAlaSerAlaAlaArgValA  437

Qy    836  ----TACACAAAGGAAGATGAGGGCTATGGCGCCCCAAAGAAGGAACCAAAACTGCTGA  890
                   |||      :::   :::
Db    437  laLeuArgThrArgGlnValAlaSerCysValAlaProLeuAspGlnLeuArgSerAlaT  457

Qy    891  AAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGT  940
               ||||||      :::
Db    457  yrValProAsnArgProSerAspThrProProGlyGlyProTrpProCys  473
```

```
RESULT 9
US-10-425-114-63655
;  Sequence 63655, Application US/10425114
;  Publication No. US2004003488A1
;  GENERAL INFORMATION:
;  APPLICANT: Liu, Jingdong
;  APPLICANT: Zhou, Yihua
;  APPLICANT: Kovalic, David K.
;  APPLICANT: Screen, Steven E
;  APPLICANT: Tabaska, Jack E
;  APPLICANT: Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53313)B
;  CURRENT APPLICATION NUMBER: US/10/425,114
;  CURRENT FILING DATE:      2003-04-28
;  NUMBER OF SEQ ID NOS: 73128
;  SEQ ID NO 63655
;  LENGTH: 93
;  TYPE: PRT
;  ORGANISM: Zea mays
;  FEATURE:
;  OTHER INFORMATION: Clone ID: LIB3153-013-B3_FLI.pep
US-10-425-114-63655
```

Alignment Scores:
Pred. No.:   0.0693
Score: 126.50
Percent Similarity: 51.61%
Best Local Similarity: 34.41%
Query Match: 6.15%
DB: 12

Length: 93
Matches: 32
Conservative: 16
Mismatches: 32
Indels: 13
Gaps: 2

US-10-644-659A-1 (1-1146) x US-10-425-114-63655 (1-93)

```
Qy  868 AAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAA 927
             ::|||       |||  :|||   |||||:|||  ::|||
Db    4 ArgAspAlaMetAsnValGluGluvalGlyGluvalGlyArgLeuLysGluGluHisArg--- 22

Qy  928 ATGATGACATGCTGCTTCATTATCTGCACAATGGCTCGCCACAGACGAGATGGCAAGATC 987
                     :::  |||       :::     ||||||
Db   23 -----------LeuGlyGlnGlnGlnProAspGlySerTyr 32

Qy  988 CAGGTTACTTTTGGAGATCTCTT-----GACAGATACGTTCGTATTCAGATAAAGTA 1041
         ::|||   ||||||           ::: :: |||
Db   33 LysValLysPheGlyValLeuPheAsnAspAspArgCysAlaAsnIlePheGluAlaLeu 52

Qy 1042 GTGGGCATTCTCATGGCGTGCCAGGAAACATGGACGTGGTAGACTTTGAAGAGAGATGCTA 1101
          |||  :||| ||  ::::: :: |||||  :|||||||||:|:|||
Db   53 ValGlyThrLeuArgAlaAlaLysLysArgIleLeuThrTyrGluGlyGluLeuLeu 72

Qy 1102 TGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
          |||:||| ||||||||||:
Db   73 LeuGlnGlyValHisAspAsnValGluIleThrLeuLeu 85
```

Alignment Scores:
Pred. No.:   0.241
Score: 126.50
Percent Similarity: 34.23%
Best Local Similarity: 27.31%
Query Match: 6.15%
DB: 15

Length: 19652
Matches: 71
Conservative: 18
Mismatches: 85
Indels: 86
Gaps: 15

US-10-644-659A-1 (1-1146) x US-10-084-846A-7 (1-19652)

```
Qy    8 CGGGCGAAAGG---------AAAGCGGGGAGGGGCCCAGCCAAGACGCGCCCTCCGGA 55
         |||||:::|||        ::|  ||||| |||||
Db 8604 ArgAlaAargArgProArgSerThrAlaAlaGlyGlyGlyArgSerArgAlaAlaSerPro 8623

Qy   56 AGATACGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGTT------GGCAGCAGT 106
```

```
         :::||||||       :::
Db 8624 ThrAlaProArgProGlyThrGlyArgSerProArgThrThrIleSerGlyAlaSer 8643

Qy  107 GGGGCAATGAGAACAGCATCAGGCAGCCCAGGAGCCTACAGGCTGGCTGC--------- 157
         :::       |||       |||||
Db 8644 TrpHisProSerArgSerProGlyArgArgThrSerArgThrAlaGlyCysArgArgCysCys 8663

Qy  158 --------------------CGGGAGGGACCCAGGAGACTCACCTCAAGCTCCTAAACCAA 196
                             |||
Db 8664 ArgThrGlyCysArgGlyArgArgGlyArgGlyThrProArgThr------ 8676

Qy  197 TCACACCCCTACTTCACACCAGAAAGCTCAGAGTG------CCCCAAAGTCGCCACCCC 250
              :::||||||    |||  :::|||
Db 8677 ------SerArgArgGlyArgValArgProAspArgArg-SerPr 8690

Qy  251 GCCTGCCAGAGGACATGGACAAGCTCAGAGAAAGC---------CCCTGAGG 301
              ::: |||||||        :::   ||||||
Db 8690 oProCysArgArgThrAlaArgCysSerAlaGlyArgArgThrSerArgProGlyProArgGl 8710

Qy  302 TTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAG 361
             ::: :::      ||| :::
Db 8710 yAspSerSerArgThrArg----------------------Th 8718

Qy  362 GAGGGGACGTGAGCCACCTCAGCCACAGTACGAGAGGGATGCTGG------ 407
          ||| ||| |||||||        |||
Db 8718 rArgAlaArgArgProAlaSerProAlaGlySerAlaAlaCysTrpGlyCysThrProAr 8738

Qy  408 ----------------------------TGT-----GCCAGAGA 433
                                        :::  ||| :
Db 8738 gSerAlaLeuProHisArgAspCysProValArgSerGlySerTrpAlaArgsSerArgArgGl 8758

Qy  434 ATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGCGGAGGAGAAAATGTGCCA 493
         ||||||       :::|||      |||
Db 8758 nArgArgArgAlaArgProAlaProProArgIleArgThrAlaArg---------CysGl 8775

Qy  494 ACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGAGCGAGGAGGAGCCCACATGGA 553
                  |||
Db 8775 y----------ThrSerArgProHisSerSerGly------ 8784

Qy  554 GGAGTGACAGCGTAGACACAGAGCACAGCGGCTATGGAGGAGGAGCGGCTGAGGAGAG----- 608
         ||||||       |||
Db 8785 ----------ArgArgHis------ArgArgTrpArgArgThrGlyGlySerThrAl 8798

Qy  609 -GCCCGAGCAGGATGGAGTGCAGGTGCAGGTGGTCAGGATCAAGCGCCCCTTGCCCTC 665
          :::||||       ::: |||
Db 8798 aAlaHisThrTrpTrpAspTrpAlaGlyAlaAlaGlyAlaGlySerTyrArgValProLeuArgLeu 8817
```

Alignment Scores:

```
Pred. No.:              0.132      Length:            645
Score:                125.50      Matches:           100
Percent Similarity:    30.99%     Conservative:      32
Best Local Similarity: 23.47%     Mismatches:        104
Query Match:           6.10%      Indels:            190
DB:                     12         Gaps:              20

US-10-644-659A-1 (1-1146) x US-10-425-114-45380 (1-645)

Qy      9 GGGCGAAAAGGAAAGCGGGAGGGCCAGCCAAGAG---CGCCCTCCGGAAGATACGGCAC 65
            |||    ::: ||||||||||||||                    :::   |||
Db     93 GlyProArgArgHisGlnGlyLysProLeuProGlnProGlyLysGlyUAlaHis 112

Qy     66 AGCCACCCTGGTCATCAGCTTGGCCCGAGTTGCCACGAGTGGCGAATGAGAACAGCAT 125
            ||| ||||||||                     ||||||||   |||:::
Db    113 GlyHisHisLeuHisArgArgGlnArgArgLeuAla--GlnGlyGln----------G 128

Qy    126 CAGGCAGGCCCAGGAGC----------------CTACAGGCTGGCTGCCGGGAG---- 163
            |||||||||||||:::               :::   |||||||
Db    128 lnGlyArgProArgAlaAlaHisArgAlaValProAspGluGlnGlyCysArgGluAlA 148

Qy    163 ------------------------------------------------------------ 163

Db    148 spProArgProHisArgArgLeuArgArgProArgProArgProLeuProGlyArgValValP 168

Qy    164 ----GGACCCAGG------------------------------ACTCACCTCAAGC 185
            ||||||||||                                   :::
Db    168 roAlaGlyProArgGluGluGlnGlyLeuGluArgArgAlaValAlaValHis--ArgSer 187

Qy    186 TCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGTTCAGAGTGCCCAAAGTCGCC 245
            :::    |||  |||    ::: |||  |||||||||::::::
Db    187 rProProAlaValArgProAlaAlaArgGlnArgGlyLysAspHisProProArgAlaPr 207

Qy    246 ACCC---------------------CGCCTGCCAGAGGACAT------GGAGATGG 275
            ||||                      |||  |||||||   |||||||
Db    207 oProGlyArgGluArgGluAspHisArgArgProAlaGlyHisArgGlnGlyAspGl 227

Qy    276 ACAAAGCTCAGAGAAAGCCCCTGAGGTTTCCACATCAAAAAGAAAGAGGTGTCCAAAAC 335
            |||||||||||                :::     |||
Db    227 y---------AlaAlaPro--GlyHisGly--GlnGlnHisValProValHisHis 241

Qy    336 ------GGTGGTCAGCAAGACTTACGAGAGAGGAGGGACCTGAGCCACCTCAGCCACAG 389
            |||    :::  ||||||||||
Db    242 ProAlaGlyIleArgGlnInAspArgArgArgGlyArg---------------- 253

Qy    390 GTACGACGAGGGATGCTGGTGTGCTTGAACCTGGCAGCAGAGAATGACATTGACAGAAT 449
            --------------------------------------------------------
Db    253 ----------------------------------------------------------- 253

Qy    450 CCTCCACCACGGCTTCCCACAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCT 509
            |||||||| |||       :::  |||
Db    254 ---ProGluHisArgArgAlaAspArgArgGluGlyArgArgAlaArgGlyValProGly 272

Qy    510 AACCAAGGGCTGAGAGTGATGGAGCAGGAGGA------GCCCACATG-------- 551
            ||||||||||              ::: |||  ||||||||||
Db    273 AlaGlnValArgAspArgArgGluAlaAlaAlaAlaGlyProGluAlaHisGluArgHisAspArg 292

Qy    552 -----------------------GAGGAGTGACAGCCTAGACACAGAGGACACGCGG 584
                                     |||||||||||||
Db    293 GlyArgArgGluArgArgAlaAlaGlyAlaGluGluGlyLysHisArgGlyHisGlyLysArg 312

Qy    585 CTATGGAGGAGGAGGGCTGAGGAGAGGCCCGAGCAGGAGTGCAGGTGGCTGTGGTCAG 644
            |||                              :::
Db    313 ----ArgAspGlyIleArgGlyProGluArgValGlyHisArgHisArgAlaAlaAspGlyAlaAlaArgAla 330

Qy    645 GATCAAGCAGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGC 704
            :::::::                               ::
Db    331 GluArgAsp-------------------------------ArgGluArgAlaAlaHisGln 340

Qy    705 CCAACAGAAATATAGCCCATGGCAACTTGAAAGGAGGATGGCAGCAGTGGGCTGATGA 764
            :::||||                        ::::::
Db    341 ProArgHisLeuProAlaHisGluGluLeuLeuHisHis------------------ 352
```

```
Qy    765 ACACATACAATCCAGAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGGCCAT 824
                                                         |||||
Db    353 ------------------------------------------LeuArgArgValHis 357

Qy    825 GTCCACCCGCCT-----------ACACAAAGGAGATGAGGGCTA-------------- 857
            ||||||
Db    358 HisHisProHisArgAlaGlyLeuArgArgAlaGlyLeuGluValArgLeuArg 377

Qy    858 --------------------------TGGCCGCCCCAAGAGAGAACCAAAACTGCTGAAAGGGC 896
                                        |||||||||||||:::.::
Db    378 AlaLeuHisGlyIleAlaHisHisCysHisProGlnArgArgHisHisHis------------ 393

Qy    897 CAAGCGCTGCTGAGGAGCACATCACAGGGA----------------------------- 926
            |||||||||||
Db    394 ----------------AspHisLeuGlnGlyProArgGluAlaValAlaAlaAspAlaArgLeu 408

Qy    927 ------------------AATGATGACATGTGCTTCATTATCTGCACAATGGCTCG 965
                                |||||||
Db    409 ValGluLeuAlaGlnGlyAspLeuProArgHisGlyHisArgAlaArgAlaArgAspLeuHisGly-AlaAr 428

Qy    966 CCAC 969
            |||
Db    428 gHis 429
```

```
RESULT 12
US-10-425-114-45434
; Sequence 45434, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
  APPLICANT:  Liu, Jingdong
  APPLICANT:  Zhou, Yihua
  APPLICANT:  Kovalic, David K.
  APPLICANT:  Screen, Steven E
  APPLICANT:  Tabaska, Jack E
  APPLICANT:  Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
  CURRENT APPLICATION NUMBER: US/10/425,114
  CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
  SEQ ID NO 45434
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700237247_FLI.pep
US-10-425-114-45434

Alignment Scores:
Pred. No.:              0.137      Length:            339
Score:                124.50      Matches:           86
Percent Similarity:    36.56%     Conservative:      31
Best Local Similarity: 26.88%     Mismatches:        106
Query Match:           6.05%      Indels:            97
DB:                     12         Gaps:              16

US-10-644-659A-1 (1-1146) x US-10-425-114-45434 (1-339)

Qy     15 AAAGGAAAGCGGGAGGGCCCAGCCAAGAGGCGCCCTTCCGGAAGATACGGCACAGCCACCCT 74
            :::.||||:||  :::|||||||||||||||:::CGGAAGATACGGCAC--------HisPro
Db     63 ArgGlyArgProGlnGluProAlaGlnGluArgProAlaProVal-----HisPro 80

Qy     75 GGTCATCAGCTTGGCCCGAGGTTGCGCAGCAGGTGGGCGAATGAGAACAGCAGGAGGC 134
            ||||||::::                    :::  |||||
Db     81 GlyHisGlnArgValPro------AlaGlyAlaGlu------HisGluGlyAla 95

Qy    135 CCAGGAGGCTACAGGGTGGCTGCGGGAGGACCCAGGACCTCACCTCAAGCTCCTAAACC 194
            ||||||||||||||||||                       :::
Db     96 AlaArgAspHisArgGlnAla-Pro-MetGluProAspHisLeuArg-------- 111

Qy    195 AATCACACACCCCTACTTCACACCAGAAAGTTCAGAGTGCCCAAAGTCGCCACCCGCCT 254
            :::::::
```

US-10-425-114-70922

Alignment Scores:
Pred. No.:   0.174   Length: 422
Score: 123.50   Matches: 85
Percent Similarity: 33.52%   Conservative: 33
Best Local Similarity: 24.15%   Mismatches: 98
Query Match: 6.00%   Indels: 138
DB: 12   Gaps: 17

US-10-644-659A-1 (1-1146) x US-10-425-114-70922 (1-422)

```
Qy      5 CTCCGGGCGAAAAGGAAAGCGGGGAG------------------------------       31
            ::|||||||||::: ||||||
Db     20 ValArgAlaArgAlaArgGlyArgArgCysLeuArgGlyGlyGluIleVal       39

Qy     32 GCCCAGCCAAGAGCGCCCTCCGAAGATACCGCACCAGCCACCTGGTCATCAGCTTGCCC   91
            |||||                             |||||||
Db     40 GlnGlnProArgValProAspValArgArgProAlaProPro-----------       53

Qy     92 GAGGTTGCCAGCAGTGGGCGAATGAGAACAGCATC-----------AGGCAGGCCCAG  138
            |||
Db     54 ------------ArgValAlaAlaLeuValLeuArgArgArgGlyVal       66

Qy    139 GAGCCTTACAGGCTGGCTGCCGGGAGGACCAGGACTCACCTTCAAGCTCCTAAACCATC  198
            |||||||
Db     67 ValProValGlyGlyValProGlyGly-----------       75

Qy    199 ACACCCCTACTTCACACCAGAAAGTCAGAGTGCCCAAAGTCGCCACCCGCCTGCCA  258
            ::: ||||||
Db     76 ---ProAlaValArgArgAlaValGlyArgLeuGlyProGlu-----ProGlnGlyPro       92

Qy    259 GAAGGACATGGAGGATGGACAAAGCGTGGTCAGGACTTACGAGAGAGGAGGGACGTGAGCCAC  318
            ::: |||||||||
Db     93 Glu---HisGlyArgHisAlaGlyAlaGlyLeuGlyArgProGlnArgArgArgArg  111

Qy    319 AAAGAGGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGAGGGACGTGAGCCAC  378
            ||||||||
Db    112 ArgArgVal----ArgGlyAlaAspLeuGlnHis  121

Qy    379 CTTGAACCTGGGCAGCCAGGAG  432
            |||||||
Db    122 LeuArgHisLeuAlaLeuProGlyGlyProProAlaProArgLeuProAspValArgProArg  141

Qy    433 AATCGACATTGACGA------------ATCCTCCAGCCACGGCCGCTCCCCAACG  474
            ::: |||||||
Db    142 ArgArgAlaArgArgProArgProAspLeuArgHis-AlaGlnAlaProGlyAr  161

Qy    475 C---GGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATG  531
            |||||||||||
Db    162 gHisGly-ArgArgArgGlyGlnProGlyVal  171

Qy    532 GAGCAGGAGGAGCC------------------------------------------------  545
            |||
Db    172 --AlaGlyGlyAlaLeuLeuArgLeuArgArgAspProAspThrArgAsnGluArgGlnHisG  191

Qy    546 ------------CACATGGAGGAGTGACAGCGTAGACACAGAGGACAGGGCGTATGGAG--  592
            ::::::
Db    192 luArgGlnGlyHisAlaGluProProHisArgArghisArgArgGlnArgArgTrpGlyGly  211

Qy    593 ------------GAGAGGCTGAGGAGGACCCGAGCAGGAGGATGGAGTGC-----------  628
            :::
Db    212 lnTyrProGlnGluGlnValLysTyrAlaArgAspProGluArgGluLeuGlyProProA  231

Qy    629 ------------AGGTGGCTGTGGTCAGGATCAAGCGCGCCCTTGCCGTCCGGTAAACAG  677
            :::::
Db    232 laAlaHisGlyArgHisLeuArgSer----------ValProGlnAspValG  245

Qy    678 ATTTACAGAGAAAACTTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAA  737
            :::::|||
Db    246 lnValArgProGlyLeuValLeuProProArgGlyVal-ProAlaGlyValArg  264

Qy    738 AGGGAGATGGCAGCAGTGGCTGATGAACACATACAATCCCAGAAGTCAATCCTTTCAG  797
```

Db   265 ArgArgGly--------------------------------- 267
Qy   798 TGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTTCACCACAAAGAGAGATGAGGG--- 854
Db   268 --------------ValArgLeuGlyHisArgHisProProArgGluProValGlnGlyAsp 283
Qy   855 ------CTATGGCGCCCCAAAGAGGAAC 878
Db   284 HisLeuLeuAlaProProGluHisArgAsp 293

RESULT 14
US-10-425-114-71620
; Sequence 71620, Application US/10425114
; Publication No. US2004003488A1
GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71620
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROB73024G01_FLI.pep
US-10-425-114-71620

Alignment Scores:
Pred. No.:   0.178
Score: 123.50
Percent Similarity: 33.52%
Best Local Similarity: 24.15%
Query Match: 6.00%
DB: 12

Length: 460
Matches: 85
Conservative: 33
Mismatches: 98
Indels: 138
Gaps: 17

US-10-644-659A-1 (1-1146) x US-10-425-114-71620 (1-460)

Qy   5 CTCCGGCGAAAAGGAAAGCGGGGAG------------------------- 31
Db   58 ValArgAlaArgAlaArgGlyArgValArgArgArgCysLeuArgGlyGlyGluIleVal 77
Qy   32 GCCCAGCCAAGAGCGCCTCCGGAAGATACGCACAGCCACCCTGGTCATCAGCTTGGCCC 91
Db   78 GlnGlnProArgValProAspValArgArgProAlaProPro------------------ 91
Qy   92 GAGGTTGGCAGCAGTGGGCGAATGAGAACAGCATC---------------AGGCAGGCCCAG 138
Db   92 ----------------ArgArgValAlaAlaLeuValLeuArgArgArgGlyVal 104
Qy   139 GAGCCTACAGGCTGGCTGCCGGGAGGGACCCAGGAGCTCACCTCAAGCTCCTAAACCAATC 198
Db   105 ValProValGlyGlyValProGlyGly-------------------- 113
Qy   199 ACACCCCTACTTCACACCAGAAAGCTTCAGAGTGCCCAAAGTCGGCCACCCGCCTGCCA 258
Db   114 ---ProAlaValArgArgValGlyArgLeuGlyProGlu------ProGlnGlyPro 130
Qy   259 GAAGGAACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGGAGGTTTCTCACATCAAAAG 318
Db   131 Glu---HisGlyArgHisAlaGlyHisAlaGlyLyArgProGlnArgArgArgArgArg 149
Qy   319 AAAGAGGTGTCCAAAACGTGGTCAGCAAGACTTACGAGAGAGGAGGACCTGAGCCAC 378
Db   150 ArgArgVal------------------------ArgGlyAlaAspLeuGlnHis 159

Qy   379 CTCCAGCCCAAGGTACGAGAGGGATGCTGGTGTG------CTTGAACCTGGGCAGCCAGAG 432
Db   160 LeuArgHisLeuAlaArgLeuProGlyGlyProProArgLeuProAspValArgProArg 179
Qy   433 AATGACATTGACAGA-----------ATCCTCCACAGCCACGGCTCCCCAACG 474
Db   180 ArgArgAlaArgArgArgProArgProAspLeuArgHis-AlaGlnAlaProGluAr 199
Qy   475 C---GGAGGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGGAGTGATG 531
Db   199 gHisGly-ArgArgArgGlyGlnProGlyVal------------------- 209
Qy   532 GAGCAGGAGGAGCC------------------------ 545
Db   210 --AlaGlyGlyAlaLeuLeuArgArgAspProAspThrArgAsnGluArgArgGlnHisG 229
Qy   546 ----CACATGGAGGAGTGACAGCGTAGACACAGAGGACCGGCTATGGAG-- 592
Db   229 luArgGlnGlyHisAlaGluProProHisArgArgHisArgArgGlnArgTrpProGlyG 249
Qy   593 --------------GAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGC---------- 628
Db   249 lnThryrProGlnGluGlulysTyrAlaArgAlaArgAspProGluArgGluGlyProProA 269
Qy   629 --------------AGTGGCTGTGTCAGGATCAAGCGCCCCTTGCCCTCCAGGTAAACAG 677
Db   269 laAlaHisGlyArgHisLeuArgSer-------------ValProGlnAspValG 283
Qy   678 ATTTACAGAGAAACTCAACTGCAAGCCCAACAGAAAATAGCCCAGTGGCAACTTGAA 737
Db   283 lnValArgArgProGlyLeuValLeuProProArgGlyVal-ProAlaGlyValArgValArg 302
Qy   738 AGGGAGATGGCAGCAGTGGGCGTGATGAACACATACAATCCCAGAGAGCTCAATCCTTTCAG 797
Db   303 ArgArgGly------------------------ 305
Qy   798 TGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTCACCACAAAGAGAGATGAGGG--- 854
Db   306 --------------ValArgLeuGlyHisArgHisProProArgGluProValGlnGlyAsp 321
Qy   855 ------CTATGGCGCCCCAAAGAGGAAC 878
Db   322 HisLeuLeuAlaProProGluHisArgAsp 331

RESULT 15
US-10-029-386-33391
; Sequence 33391, Application US/10029386
; Publication No. US20030194704A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 33391
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002104.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
US-10-029-386-33391

Alignment Scores:

us-10-644-659a-1.n2p.rapb

Thu May 6 09:39:11 2004

Search completed: May 4, 2004, 08:59:57
Job time : 129.5 secs

```
Pred. No.:            0.213       Length:          293
Score:               122.00      Matches:          70
Percent Similarity:  28.15%      Conservative:     15
Best Local Similarity: 23.18%    Mismatches:       99
Query Match:          5.93%      Indels:          118
DB:                              Gaps:             12

US-10-644-659A-1 (1-1146) x US-10-029-386-33391 (1-293)

Qy      62 GCACAGCCACCCGGTCATCAGCT--------------------------------- 85
              ||||::| |||||| ||||:::
Db       7 AlaLysTrpProTrpGlySerArgArgLeuAspProGlyAlaArgThrGlyAlaLys 26

Qy      86 TGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAACAGCATCAGGCAGG-------- 133
              ||||||||||| |||||| |||| ||| ||| |||| |||||
Db      27 TrpProGlySerGlySerGlyLysArgArgProAlaArgGlyArgCysProGlyPro 46

Qy     134 ----------------------------------------CCCAGGAGCCTACAGGCT 151
                                                      |||:::|||||||
Db      47 ThrGlnLeuProProAlaProArgAlaProArgProSerArgAlaLeuGlnPro 66

Qy     152 GGCTGCCGGGAGGGACCCCAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTT 211
              |||||||                        ::: |||||
Db      67 ------ArgProGlyPro--------------------AlaGlnArgProArgProLeuPhe 78

Qy     212 CACACCAGAAAGTCAGA----------------------GTGCCCAAAGTCGCCACCCC 250
              |||||                                   |||:::
Db      79 HisArgArgProAlaArgArgGluAlaGlnGlnGlyCysAspProArgAlaHisPro 98

Qy     251 GCCTGCCAGAAGGACAATGGAGATG-----------------GACAAAGCTCAGAGAAAGCCC 295
              |||||                          |||:::|||
Db      99 AlaArgGlnAlaProTrpGluMetArgGlyHisArgGlnGlyProAlaGluArgGlnAlaGln 118

Qy     296 CTGAGGTTTCTCACATCAAAAAGAAAGAGTGTCCAAAACGGTGGTCAGCAAGACTTACG 355
              ||| |||               |||:::: ||| :||
Db     119 ArgArgSerLeuProHisProArgProArgProAlaProArgAsnArgArgThrArgAlaVal 138

Qy     356 AGAGAGGAGGGAGGTGAGCCACCTCAGCCACAGTACGAGAGGGATGCTGGTGTGCTTG 415
              |||||||||              ||| |||||| |||
Db     139 ProAlaGlyGlyGlyGlyAlaThrGlnAlaThrGlnAlaGluGlyThrLeuAlaAsnPheAlaProPro 158

Qy     416 AACCTGGGCAGCCAG--------------------------------------- 430
              :::||||||:::
Db     159 GluValAlaSerArgArgProGlyAlaHisProProAlaAlaArgProArgProGlyAla 178

Qy     431 ----------------------------------------AGAATGACATTGACA 445
                                                      |||:::
Db     179 ArgProProGlnProGlyArgLeuGlyArgArgArgCysAspLeuArgValGlnSerAla 198

Qy     446 GAATCCTCCACAGCCACGGCTCCCCAACCGGA------GGAGAAAATGTGCCAACCTGG 499
              :::                             ||| |||||
Db     199 ArgArgAlaAlaArgThrProProProArgGlyProProAlaArgArgAlaLeuThrTrp 218

Qy     500 ------------------------------------------------TGT 502
                                                              |||
Db     219 AspIleCysGlyArgLysLeuMetSerLeuArgSerMetGluProGlyGluArgLeuCys 238

Qy     503 CTGAGCTAACCAAGGCTGGAGAGTGATGAGGAGCAGGAGGCCACCATGAGGAGTGACA 562
              ||| |||||||                            ||||:::||
Db     239 SerSer----------------------TrpHisArgThrProSerArgSerAlaCys 251

Qy     563 GCGTAGACACAGAGAGAGACAGCGGCTATGGAG-----------------GAGAGGCTG 601
              |||  ::::::  |||||||
Db     252 AlaArgSerLysThrThrAlaGluHisThrArgTrpLeuGlyArgThrLeu 271

Qy     602 AGGAGAGGGCCCGAGCAGCAGGATGGAGTGGAGTGGCTGTGGTCAGGATCAAGCGCCCCTTGC 661
              :::                                 ||||||
Db     272 GlnSerMetSerGlnArgProHisCysArgHisCys------SerMetAlaProAla 288

Qy     662 CCTCCC 667
              |||
Db     289 GlyPro 290
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:   May  4, 2004, 08:49:07 ; Search time 64 Seconds
                                   (without alignments)
                          10118.733 Million cell updates/sec

Title:          US-10-644-659A-1
Perfect score:  2058
Sequence:       1 atggcccgggcgaaagga.........tgattacgctactcaagtga 1146

Scoring table:  BLOSUM62
                Xgapop   10.0 ,  Xgapext  0.5
                Ygapop   10.0 ,  Ygapext  0.5
                Fgapop    6.0 ,  Fgapext  7.0
                Delop     6.0 ,  Delext   7.0

Searched:       1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:       3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10644659/runat_04052004_084901_28500/app_query.fasta_1.1287
-DB=A_Geneseq_29Jan04 -QFMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CGN 1_81 @runat_04052004_084901_28500 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :    A_Geneseq_29Jan04:*
        1: geneseqp1980s:*
        2: geneseqp1990s:*
        3: geneseqp2000s:*
        4: geneseqp2001s:*
        5: geneseqp2002s:*
        6: geneseqp2003aa:*
        7: geneseqp2003bs:*
        8: geneseqp2004s:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2029	98.6	381	7	ADB65015	Adb65015 Human pro
2	1988	96.6	456	4	ABG27995	Abg27995 Novel hum
3	647	31.4	121	4	AAM21624	Aam21624 Peptide #
4	647	31.4	121	4	ABB43984	Abb43984 Peptide #
5	647	31.4	121	4	AAM37926	Aam37926 Peptide #
6	647	31.4	121	4	ABB26893	Abb26893 Protein #
7	647	31.4	121	4	AAM77709	Aam77709 Human bon
8	647	31.4	121	4	AAM64987	Aam64987 Human bra
9	647	31.4	121	4	ABG59364	Abg59364 Human liv
10	362	17.6	69	3	AAG00337	Aag00337 Human sec
11		17.6	362	2	AAY11623	Aay11623 Human 5'
12		12.5	258	4	ABB60049	Abb60049 Drosophil
13		10.0	205	4	ABB58876	Abb58876 Drosophil
14		9.2	190	4	AAM21625	Aam21625 Peptide #
15		9.2	190	4	ABB43985	Abb43985 Peptide #
16		9.2	190	4	AAM37927	Aam37927 Peptide #
17		9.2	190	4	ABB26894	Abb26894 Protein #
18		9.2	190	4	AAM77710	Aam77710 Human bon
19		9.2	190	4	AAM64988	Aam64988 Human bra
20		9.2	190	4	ABG59365	Abg59365 Human liv
21		6.6	136	6	ABB98398	Abb98398 Streptomy
22		6.5	134	6	ABP76682	Abp76682 Streptomy
23		6.5	133.5	3	AAB32055	Aab32055 Human sec
24		6.5	133.5	3	AAG58353	Aag58353 Arabidops
25		6.4	132	6	ABO53053	Abo53053 Human put
26		6.4	131.5	3	AAG58354	Aag58354 Arabidops
27		6.3	130.5	6	AAE36105	Aae36105 Human chr
28		6.3	130.5	6	AAE36106	Aae36106 Human chr
29		6.2	128.5	4	AAB59813	Aab59813 TutD prot
30		6.2	128.5	4	AAB59826	Aab59826 Protein #
31		6.2	128	6	ABP76680	Abp76680 Streptomy
32		6.2	127	6	ABP76679	Abp76679 Streptomy
33		6.1	125.5	6	ABG30256	Abg30256 Novel hum
34		6.0	124.5	7	AAG58854	Aag58854 Zea mays
35		6.0	123	4	ABB61050	Abb61050 Drosophil
36		5.9	121.5	4	ABG05146	Abg05146 Novel hum
37		5.9	121	4	ABB65135	Abb65135 Drosophil
38		5.9	121	5	AAE19798	Aae19798 Mouse Hai
39		5.9	121	4	ABU62544	Abu62544 Mouse Hai
40		5.9	120.5	4	ABB11928	Abb11928 Human BG3
41		5.9	120.5	4	ABG06417	Abg06417 Novel hum
42		5.9	120.5	4	ABG02138	Abg02138 Novel hum
43		5.9	120.5	5	ABB08161	Abb08161 Human cyt
44		5.8	120	5	ABB79905	Abb79905 Activator
45		5.8	120	4	ABG03533	Abg03533 Novel hum

```
                              ALIGNMENTS

RESULT 1
ADB65015
ID   ADB65015 standard; protein; 381 AA.
XX
AC   ADB65015;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human protein encoded by clone SKMUS20026340.
XX
KW   Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
KW   cell regeneration; membrane protein; signal transduction-related protein;
KW   transcription-related protein; osteoporosis; neurological disease;
KW   cancer; tumour.
XX
OS   Homo sapiens.
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki M, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-450961/43.
```

DR   N-PSDB; ADB63045.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein. Note: Some of the
CC   sequence presented is a protein of the invention. Note: The printed
CC   sequence data for this patent is not represented in the printed
CC   specification, but is based on sequence information supplied by the
CC   European Patent Office.
XX
SQ   Sequence 381 AA;

Alignment Scores:
Pred. No.:   5.25e-181   Length: 381
Score: 2029.00   Matches: 381
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 98.59%   Indels: 0
DB: 7   Gaps: 0

US-10-644-659A-1 (1-1146) x ADB65015 (1-381)

Qy   1 ATGGCTCCGGCGAAAAGGAAAGCGGGAGGGCCCAGCCAAGAGGGCCCTCCCAAGAGATA   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaProGlyLysGluLysGlySerGlyGlyProAlaLysSerAlaLeuArgLysIle   20

Qy   61 CGCACAGCCACCCCTGGTCATCAGGTTGGCCCGAGGTTGGCAGCAGTGGCGAATGAGAAC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpArgAsnGluAsn   40

Qy   121 AGCATCAGGCAGGCAGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCT 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyThrGlnAspSerPro   60

Qy   181 CAAGCTCCTAAACCAATCACACCCCTACTTCACACAGAAAGTCAGAGTGCCCAAAG 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GlnAlaProLysProIleThrProThrSerHisSerGlnLysSerAlaProLys   80

Qy   241 TCGCCACCCCGCCTGCCAGAAGGACATGAGAATGGCAAGGCCCGAGATGACCATGTTGTGATTACGCTACTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerProProArgLeuProGluGlyHisGluMetAlaArgProGluMetThrMetLeu 100

Qy   301 GTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGTCAGCAAGACTTACGAGAGA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ValSerHisIleLysLysGluValSerLysThrValSerLysThrTyrGluArg 120

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTAGCAGGATGCTGGTGTGTGCTTGAACCT 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GlyGlyAspValSerHisLeuSerHisArgTyrArgGluTyrArgAspHisValLeuGluPro 140

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCAGCACGCTGCCCCAACGCCGGAGG 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 GlyGlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisGlySerProThrArgArg 160

Qy   481 AGAAAATGTGCCAACTGGTCTGTCTGAGCTACCAAGGCTGGAGAGTGATGGAGCAGGAG 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu 180

Qy   541 GAGCCCACATGGAGGAGTGACACGCTGAGACACAGAGACACGGGCTATGGAGGAGGAGGCT 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GluProThrTrpArgSerAspSerValAspThrGluArgGlyTyrGlyGlyGlyGlyAla 200

Qy   601 GAGGAGAGGCCCAGACAGGATGGAGTGCAGGTGCAGGATCAGCAACGCGCCCCTTG 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 GluGluArgProGlnGlnAspGlyValGlnValAlaValAlaArgIleLysArgProLeu 220

Qy   661 CCCTCCCAGGTAAACAAGATTTACAGAGAAAACTCAACTGACAAAGCCCAACAGAAATATAGC 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ProSerGlnValAsnArgPheThrGluPheThrLysLeuAsnCysLysAlaGlnGlnLysTyrSer 240

Qy   721 CCAGTGGCAACTTGAAAGGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAG 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ProValGlyAsnLeuLysGlyArgTrpGlnTrpAlaAspGluHisIleGlnSerGln 260

Qy   781 AAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 LysLeuAsnProPheSerGluGluPheAspTyrGluLeuAlaMetSerThrArgLeuHis 280

Qy   841 AAAGGAGGATGAGGGCTATGGCCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 LysGlyAspGluGlyTyrGlyArgProLysGluGluThrLysThrAlaGluArgAlaLys 300

Qy   901 CGTGCTGAGGAGGCACATCTACAGGGAAAATGATGGAAATGATGGAAMetCysPheIleIleCysThrMet 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ArgAlaGluGluHisIleTyrArgGluMetMetAspMetCysPheIleIleCysThrMet 320

Qy   961 GCTCGCCAGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 AlaArgHisArgArgAspGlyLysIleGlnValThrPheGlyAspIleLeuPheAspArgTyr 340

Qy   1021 GTTCGTATTTCAGATAAAGTAGTAGTGGCCATTCTCATCGTGCCCAGGAAAACATGGACTGGTA 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   341 ValArgIleSerAspLysValValGlyValLeuGlyIleLeuMetArgAlaArgLysHisGlyLeuVal 360

Qy   1081 GACTTTGAAGGAGGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGATTACGCTACTC 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AspPheGluGlyGlyGluMetLeuTrpGlnGlyValArgAspHisValValIleThrLeuLeu 380

Qy   1141 AAG 1143
          |||
Db   381 Lys 381

RESULT 2
ABG27995
ID   ABG27995 standard; protein; 456 AA.
XX
AC   ABG27995;
XX
DT   18-FEB-2002   (first entry)
XX
DE   Novel human diagnostic protein #27986.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX

PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92182.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 58354; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have application in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 456 AA;

Alignment Scores:
Pred. No.:   3.9e-177   Length: 456
Score: 1988.00   Matches: 381
Percent Similarity: 86.00%   Conservative: 0
Best Local Similarity: 86.60%   Mismatches: 0
Query Match: 96.60%   Indels: 62
DB: 4   Gaps: 1

US-10-644-659A-1 (1-1146) x ABG27995 (1-456)

Qy   1 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGCCAGCCAAGAGCGCCTCCGGAAGATA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14 MetAlaProGlyGluLysGluSerGlyGlyAlaGlyGlnGlyProAlaLysSerAlaLeuArgLysIle 33

Qy   61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTTGGGCGAATCAGAAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   34 ArgThrAlaThrLeuVallleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn 53

Qy   121 AGCATCAGGCAGGCCAGGAGCCTACAGGCTGCCGGGAGGGACCAGGACTCACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   54 SerIleArgGlnAlaGlnArgLeuGlnProThrGlyTrpLeuProGlyLysTrpGlnThrGlnAspSerPro 73

Qy   181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAAGCTCAGAGTGCCCCAAAG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   74 GlnAlaProLysProIleThrProProThrSerHisGlnLysLysAlaGlnSerAlaProLys 93

Qy   241 TCGCCACCCGCCTGCCAGAAGGACATGGAAGGCATGAGCAGAAAGCTCGAGAAAAGCCCCTGAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   94 SerProProArgLeuProGluProGluGluProGluHisGlyAspGlyHisGlySerSerGlyLysAlaProGlu 113

Qy   301 GTTTCTCACATCAAAAAGAAGAAGAGTGTCCAAAACGGTGGTCAGCAGACTTACGAGAGA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   114 ValSerHisIleLysLysLysGluValSerLysThrValValSerLysThrTyrGluArg 133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361 GGAGGGACGTGAGCCACCTCAGCCACCAGGTACGAGAGGATGCTGGTGTGCTTGAACCT 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134 GlyGlyAspValSerHisLeuSerHisArgTyrGluArgAspAlaGlyValLeuGluPro 153

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACGGCGGAGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   154 GlyGlnProGluAsnAspIleAspArgIleLeuHisSerHisGlySerProThrArgArg 173

Qy   481 AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 ArgLysCysAlaAsnLeuValSerGluLeuThrLysGlyTrpArgValMetGluGlnGlu 193

Qy   541 GAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGACACCGGCTATGGAGGAGAGGCT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 GluProThrTrpArgSerAspSerValAspThrGluAspThrProAlaMetGluGluArgGlyGlyGluAla 213

Qy   601 GAGGAGAGGCCCGAGCAGGATGAGTGCAGGATGCAGGTGGCTGTGGTCAGGATCAAGGCCCCTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   214 GluGluArgProGluGlnAspGluCysArgMetGlnValAlaValAlaValArgIleLysArgProLeu 233

Qy   661 CCCTCCCAG--- 669
          ||||||||
Db   234 ProSerGlnGlyGlyCysLeuHisGluGluLysTyrSerLeuTyrArgLysIleTyr 253

Qy   669 --- 669
Db   254 GlyValGlySerProPheSerLeuSerProTyrArgSerGluThrGluGluLys 273

Qy   669 --- 669
Db   274 IleMetProLeuGlyValProCysIleLeuLysLeuProLeuGlyAsnGlnIleSerGly 293

Qy   670 -------------GTAAACAGATTTACAGAGAAACTCAACTGCAAGCCCAACAGAAA 714
                       |||||||||||||||||||||||||||||||||||||
Db   294 SerGluPheGlyTrpValAsnArgPheThrGlyLeuAsnCysLysAlaGlnAsnLys 313

Qy   715 TATAGCCCAGTGGCAACTTGAAAGGGAGGATGCAGCAGTGGGCTGATGAACACTACAA 774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 TyrSerProValGlyAsnLeuLysGlyArgTrpGlnGlnTrpAlaAspGluHisIleGln 333

Qy   775 TCCCAGAAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGC 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   334 SerGlnLysLeuAsnProPheSerGluGluPheAspTyrGluLeuAlaMetSerThrArg 353

Qy   835 CTACAAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   354 LeuHisLysGlyAspGluTyrGlyArgProLysGluGlyThrLysThrAlaGluArg 373

Qy   895 GCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGC 954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 AlaLysArgAlaGluGluHisIleTyrArgGluMetMetAspMetCysPheIleIleCys 393

Qy   955 ACAATGGCTCGCCACCAGACGAGATGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGAC 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   394 ThrMetAlaArgHisArgArgAspArgLysIleGlnValThrPheGlyAspLeuPheAsp 413

Qy   1015 AGATACGTTCGTATTTCAGATAAAGTAGTGGGGCATTCTCATGGCGTGCCAGAAACATGGA 1074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   414 ArgTyrValArgIleSerAspLysValValGlyHisSerHisGlyValProGluThrTrpHisGlyVal 433

Qy   1075 CTGGTAGACTTGAAGGAGGAGATGCTATGGCCAGGCCGGAGATGACCATGTTGTGATTACG 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434 LeuValAspPheGluGlyGluMetLeuTrpProGlnGlyArgArgAspAspHisValIleThr 453

Qy   1135 CTACTCAAG 1143
          ||||||||
Db   454 LeuLeuLys 456

RESULT 3
AAM21624
ID   AAM21624 standard; protein; 121 AA.

```
XX
AC   AAM21624;
XX
DT   12-OCT-2001   (first entry)
XX
DE   Peptide #8058 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 27; SEQ ID NO 26450; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI284459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 121 AA;

Alignment Scores:
Pred. No.:                 1.04e-51          Length:          121
Score:                     647.00            Matches:         121
Percent Similarity:        100.00%           Conservative:      0
Best Local Similarity:     100.00%           Mismatches:        0
Query Match:               31.44%            Indels:            0
                                             Gaps:              0

US-10-644-659A-1 (1-1146) x AAM21624 (1-121)

Qy  262 GGACATGGAGGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCCACATCAAAAGAAA 321
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys 20

Qy  322 GAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGGACGTGAGCCACCTC 381
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40

Qy  382 AGCCACAGGTACCAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAATGACATT 441
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy  442 GACAGAATCCTCCACAGCCACGGCTCCCCACGCGGAGCGGAGAAAAATGTGCCAACCTGGTG 501
```

```
Db   61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgLysCysAlaAsnLeuVal 80
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  502 TCTGAGCTAACCAAGGCTGGAGAGTGATGAGCAGGAGGAGCCACATGGAGGAGTGAC 561
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGluProThrTrpArgSerAsp 100
            |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  562 AGCGTAGACACAGAGGACAGCGGGCTATGAGCAGGAGGGCTGAGGAGGAGCCCGAGCAGGAT 621
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 SerValAspThrGluAspSerGlyTyrGlyGlyGluAlaGluGluArgProGluGlnAsp 120

Qy  622 GGA 624
        |||
Db  121 Gly 121

RESULT 4
ABB43984
ID   ABB43984 standard; peptide; 121 AA.
XX
AC   ABB43984;
XX
DT   04-FEB-2002   (first entry)
XX
DE   Peptide #11490 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 36619; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 121 AA;

Alignment Scores:
Pred. No.:                 1.04e-51          Length:          121
Score:                     647.00            Matches:         121
Percent Similarity:        100.00%           Conservative:      0
Best Local Similarity:     100.00%           Mismatches:        0
Query Match:               31.44%            Indels:            0
DB:                                          Gaps:              0
```

US-10-644-659A-1 (1-1146) x ABB43984 (1-121)

```
Qy  262  GGACATGGAGGATGGACAAAGCTCAGAGAAGCCCCTGAGGTTTCTCACATCAAAAGAAA  321
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys  20

Qy  322  GAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGACGTGAGCCACCTC  381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21  GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu  40

Qy  382  AGCCACAGGTACGAGAGGGATGTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATT  441
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41  SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle  60

Qy  442  GACAGAATCCTCCAGCCACGCCTCCCAACGCGGAGAGAAAATGTGCCAACCTGGTG  501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal  80

Qy  502  TCTGAGCTAACCAAGGGCTGGAGAGTGATGAGCGAAGAGTGATGAGCGAGGAGTGAC  561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81  SerGluLeuThrLysGlyTrpArgValMetGluGlnGluGluArgSerTrpArgSerAsp  100

Qy  562  AGCGTAGACACAGAGACCAGAGCAGCGCGCTATGAGGAGGAGGCTGAGGAGGAGGCCCGAGCAGGAT  621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101  SerValAspThrGluAspSerGlyTyrGlyGlyLysGluAlaGluGluArgProGluGlnAsp  120

Qy  622  GGA  624
              |||
Db  121  Gly  121
```

RESULT 5
AAM37926
ID AAM37926 standard; protein; 121 AA.
XX
AC AAM37926;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #11963 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207455P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-0063236B.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0002426B.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
PS Claim 27; SEQ ID NO 38195; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AAI55466). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for

CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
SQ Sequence 121 AA;

Alignment Scores:
Pred. No.:   1.04e-51   Length: 121
Score: 647.00   Matches: 121
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 31.44%   Indels: 0
DB: 4   Gaps: 0

US-10-644-659A-1 (1-1146) x AAM37926 (1-121)

```
Qy  262  GGACATGGAGGATGGACAAAGCTCAGAGAAGCCCCTGAGGTTTCTCACATCAAAAGAAA  321
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys  20

Qy  322  GAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGACGTGAGCCACCTC  381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21  GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu  40

Qy  382  AGCCACAGGTACGAGAGGGATGTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATT  441
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41  SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle  60

Qy  442  GACAGAATCCTCCAGCCACGCCTCCCAACGCGGAGAGAAAATGTGCCAACCTGGTG  501
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  AspArgIleLeuHisSerHisGlySerProThrArgArgLysCysAlaAsnLeuVal  80

Qy  502  TCTGAGCTAACCAAGGGCTGGAGAGTGATGAGCGAAGAGTGATGAGCGAGGAGTGAC  561
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81  SerGluLeuThrLysGlyTrpArgValMetGluGlnGluGluArgSerTrpArgSerAsp  100

Qy  562  AGCGTAGACACAGAGACCAGAGCAGCGCGCTATGAGGAGGAGGCTGAGGAGGAGGCCCGAGCAGGAT  621
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101  SerValAspThrGluAspSerGlyTyrGlyGlyLysGluAlaGluGluArgProGluGlnAsp  120

Qy  622  GGA  624
              |||
Db  121  Gly  121
```

RESULT 6
ABB26893
ID ABB26893 standard; protein; 121 AA.
XX
AC ABB26893;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #8892 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000666.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-0063236B.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0002426B.

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 38015.
XX
PA
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO20015727276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO 38015; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
SQ Sequence 121 AA;

Alignment Scores:
Pred. No.:   1.04e-51   Length: 121
Score: 647.00   Matches: 121
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 31.44%   Indels: 0
DB: 4   Gaps: 0

US-10-644-659A-1 (1-1146) x AAM77709 (1-121)

Qy 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAA 321
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGlyHisIleLeuLysLysLys 20

Qy 322 GAGGTGTCCAAAACCGTGGTCAGCAAGACTTACGAGAGGAGGGGACTTGAGCCACCTC 381
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40

Qy 382 AGCCACCAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATT 441
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy 442 GACAGAATCCTCCCAGCACACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTG 501
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AspArgIleLeuHisSerHisSerGlySerProThrArgArgArgLysCysAlaAsnLeuVal 80

Qy 502 TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAAGACTCCACATGGAGGAGGTGAC 561
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerGluLeuThrLysGlyTrpArgValMetGluGlnAspSerThrTrpArgArgSerAsp 100

Qy 562 AGCGTAGACACAGAGACAGCGGCTATGGAGCAGGAGGCCCGAGGAGGCGAGCCCGAGCCAGGAT 621
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 621

XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX
PS Claim 15; SEQ ID NO 28663; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 121 AA;

Alignment Scores:
Pred. No.:   1.04e-51   Length: 121
Score: 647.00   Matches: 121
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 31.44%   Indels: 0
DB: 4   Gaps: 0

US-10-644-659A-1 (1-1146) x ABB26893 (1-121)

Qy 262 GGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAA 321
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGlyHisIleLeuLysLysLys 20

Qy 322 GAGGTGTCCAAAACCGTGGTCAGCAAGACTTACGAGAGGAGGGGACTTGAGCCACCTC 381
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40

Qy 382 AGCCACCAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATT 441
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy 442 GACAGAATCCTCCCAGCACACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTG 501
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AspArgIleLeuHisSerHisSerGlySerProThrArgArgArgLysCysAlaAsnLeuVal 80

Qy 502 TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAAGACTCCACATGGAGGAGGTGAC 561
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerGluLeuThrLysGlyTrpArgValMetGluGlnAspSerThrTrpArgArgSerAsp 100

Qy 562 AGCGTAGACACAGAGACAGCGGCTATGGAGCAGGAGGCCCGAGGAGGCGAGCCCGAGCCAGGAT 621
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 SerValAspThrGluAspSerGlyTyrGlyGlyGluAlaGluArgProGluGlnIleAsp 120

Qy 622 GGA 624
        |||
Db 121 Gly 121

RESULT 7
AAM77709
ID AAM77709 standard; protein; 121 AA.
XX
AC AAM77709;
XX
DT 06-NOV-2001 (first entry)
XX

```
Db    101 SerValAspThrGluGlyAspSerGlyTyrGlyGlyGluAlaGluGluArgProGluGlnAsp 120

Qy    622 GGA 624
          |||
Db    121 Gly 121

RESULT 8
AAM64987
ID   AAM64987 standard; protein; 121 AA.
XX
AC   AAM64987;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human brain expressed single exon probe encoded protein SEQ ID NO: 37092.
XX
KW   Human; brain expressed exon; gene expression analysis; probe; microarray;
KW   Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000667.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains.
XX
PS   Example 4; SEQ ID NO 37092; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system
CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is a protein encoded by one of
CC   the probes of the invention
XX
SQ   Sequence 121 AA;
```

```
Alignment Scores:
Pred. No.:             1.04e-51          Length:         121
Score:                 647.00            Matches:        121
Percent Similarity:    100.00%           Conservative:   0
Best Local Similarity: 100.00%           Mismatches:     0
Query Match:           31.44%            Indels:         0
DB:                    4                 Gaps:           0

US-10-644-659A-1 (1-1146) x AAM64987 (1-121)

Qy    262 GGACATGGAGGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAA 321
              |||||||||||||||||||||||||||||||||||||||||||||
Db      1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys 20

Qy    322 GAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGGACGTGAGCCACCTC 381
              ||||||||||||||||||||||||||||||||||||||||||
Db     21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40
```

```
Qy    382 AGCCACAGGTACGAGAGGGGATGGCTGCTGTGTGCTTGAACCTGGCAGCCAGAGAATGACATT 441
              ||||||||||||||||||||||||||||||||||||||||||
Db     41 SerHisArgTyrGluArgAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy    442 GACAGAATCCTCCACAGCCACGGCTCCCACGCCGGAGGAGAAAAATGCCAACCTGGTG 501
              ||||||||||||||||||||||||||||||||||||||||||
Db     61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgLysCysAlaAsnLeuVal 80

Qy    502 TCTGAGCTAACCAAGGCTGGAGAGTGATGAGCAGCAGGAGGAGCCCACATGGAGGAGTGAC 561
              ||||||||||||||||||||||||||||||||||||||||||
Db     81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGluGluProThrTrpArgSerAsp 100

Qy    562 AGCGTAGACACAGAGGACCAGCGCTATGGAGGAGGAGGCTGAGGAGAGGCCCAGCAGGAT 621
              ||||||||||||||||||||||||||||||||||||||||||
Db    101 SerValAspThrGluGluAspSerGlyTyrGlyGlyGluAlaGluGluArgProGluGlnAsp 120

Qy    622 GGA 624
          |||
Db    121 Gly 121

RESULT 9
ABG59364
ID   ABG59364 standard; peptide; 121 AA.
XX
AC   ABG59364;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human liver peptide, SEQ ID No 38012.
XX
KW   Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW   hypercholesterolaemia; coronary heart disease.
XX
OS   Homo sapiens.
XX
PN   WO200157273-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000664.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488898/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human adult liver.
XX
PS   Claim 27; SEQ ID NO 38012; 658pp; English.
XX
CC   The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC   measuring human gene expression in a sample derived from human adult
CC   liver, comprising one of 13109 defined nucleotide sequences given in the
CC   specification (or complements/ fragments). The probe hybridises at high
CC   stringency to a nucleic acid molecule expressed in the human adult liver.
CC   (I) may be used for predicting, measuring and displaying gene expression
CC   in samples derived from human adult liver. The genes identified may be
CC   involved in genetic liver diseases such as cirrhosis,
CC   hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC   associated with coronary heart disease. ABG47348-ABG59930 represent human
CC   liver single exon encoded peptides of the invention. Note: The sequence
CC   information for this patent does not appear in the printed specification
```

CC   but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences

XX

SQ   Sequence 121 AA;

Alignment Scores:
Pred. No.:   1.04e-51   Length: 121
Score: 647.00   Matches: 121
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 31.44%   Indels: 0
DB: 4   Gaps: 0

US-10-644-659A-1 (1-1146) x ABG59364 (1-121)

Qy   262 GGACATGGAGAATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTCTCACATCAAAAAGAAA 321
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GlyHisGlyAspGlyGlnSerSerGluLysAlaProGluValSerHisIleLysLysLys 20

Qy   322 GAGGTGTCCAAAAACGGTGTCAGCAAGACTTACGAGAGAGGGACGTGAGCCACCTC 381
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 GluValSerLysThrValValSerLysThrTyrGluArgGlyGlyAspValSerHisLeu 40

Qy   382 AGCCACAGGTACGAGAGAGGGATGTGGGTGCTGAACCTGGCAGCCAGAGAATGACATT 441
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerHisArgTyrGluArgGlyAspAlaGlyValLeuGluProGlyGlnProGluAsnAspIle 60

Qy   442 GACAGAAATCCTCCACAGCCACGGCTCCCCAACGCGGAGAGAAAATGCCAACCTGGTG 501
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AspArgIleLeuHisSerHisGlySerProThrArgArgArgArgLysCysAlaAsnLeuVal 80

Qy   502 TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGCAGCCCCATGGAGGAGTGAC 561
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 SerGluLeuThrLysGlyTrpArgValMetGluGlnGlnProThrTrpArgSerAsp 100

Qy   562 AGCCTAGAACACAGAGACACAGGCGGCTATCGAGGAGGAGGCCCCGAGGAGCAGGAT 621
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 SerValAspThrGluAspSerGlyTyrArgGlyGlyGlyAlaGluGluArgProGluGlnAsp 120

Qy   622 GGA 624
          |||
Db   121 Gly 121

RESULT 10
AAG00337
ID   AAG00337 standard; protein; 69 AA.
XX
AC   AAG00337;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein, SEQ ID NO: 4418.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999; 99US-0122487P.
XX
PA   (GEST) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   N-PSDB; AAC00343.
XX

PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX

PS   Claim 13; SEQ ID NO 4418; 71pp + Sequence Listing; English.

XX

CC   The present sequence is a polypeptide encoded by one of a large number of
CC   5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC   prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC   tissues. EST sequences usually correspond mainly to the 3' untranslated
CC   region (UTR) of the mRNA because they are often obtained from oligo-dT
CC   primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC   sequences derived from the 5' ends of mRNAs and even in those cases where
CC   longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC   included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC   therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC   are also used in diagnostic, forensic, gene therapy and chromosome
CC   mapping procedures. They are used to obtain upstream regulatory sequences
CC   and to design expression and secretion vectors

XX

SQ   Sequence 69 AA;

Alignment Scores:
Pred. No.:   4.25e-25   Length: 69
Score: 362.00   Matches: 68
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 17.59%   Indels: 0
DB: 3   Gaps: 0

US-10-644-659A-1 (1-1146) x AAG00337 (1-69)

Qy   1 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGCGGGCCCTCCGGAAGATA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaProGlyGlyLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle 20

Qy   61 CGCACAGCCACCCTGTCATCAGCTTGGCCTCAGCAGTGGGCAGTGGGCGAATGAGAAC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpProAlaAsnGluAsn 40

Qy   121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGAGGACCCAGGACTTCACCT 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyProGlyThrArgGlnAspSerPro 60

Qy   181 CAAGCTCCTAAACCAATCACACCC 204
          ||||||||||||||||||||||||
Db   61 GlnAlaProLysProIleThrPro 68

RESULT 11
AAY11623
ID   AAY11623 standard; protein; 70 AA.
XX
AC   AAY11623;
XX
DT   16-JUN-1999 (first entry)
XX
DE   Human 5' EST secreted protein SEQ ID NO:275.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS   Homo sapiens.
XX
PN   WO9906439-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998; 98WO-IB001233.
XX

```
PR   01-AUG-1997;    97US-00904468.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
DR   WPI; 1999-153700/13.
DR   N-PSDB; AAX40341.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from cDNA
PT   libraries derived from liver, lung, large intestine, colon, thyroid and
PT   pancreas tissue.
XX
PS   Claim 27; Page 362; 398pp; English.
XX
CC   AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY11533 to
CC   AAY11679, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell
XX
SQ   Sequence 70 AA;

Alignment Scores:
Pred. No.:              4.28e-25        Length:           70
Score:                  362.00          Matches:          68
Percent Similarity:     100.00%         Conservative:     0
Best Local Similarity:  100.00%         Mismatches:       0
Query Match:            17.59%          Indels:           0
DB:                     2               Gaps:             0

US-10-644-659A-1 (1-1146) x AAY11623 (1-70)

Qy    1 ATGGCTCCGGGCGAAAAGGAAAGCGGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATA 60
        | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MetAlaProGlyGluLysGluSerGlyGluGlyProAlaLysSerAlaLeuArgLysIle 20

Qy   61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATAGAGAAC 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ArgThrAlaThrLeuValIleSerLeuAlaArgGlyTrpGlnGlnTrpAlaAsnGluAsn 40

Qy  121 AGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGAGGACCACCAGGACTCACCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerIleArgGlnAlaGlnGluProThrGlyTrpLeuProGlyGlyThrGlnAspSerPro 60

Qy  181 CAAGCTCCTAAACCAATCACACCC 204
        ||||||||||||||||
Db   61 GlnAlaProLysProIleThrPro 68

RESULT 12
ABB60049
ID   ABB60049 standard; protein; 399 AA.
XX
AC   ABB60049;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 6939.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
```

```
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0061415O.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL04152.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 6939; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL6175) and the encoded proteins (ABB5737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 399 AA;

Alignment Scores:
Pred. No.:              4.62e-15        Length:           399
Score:                  258.00          Matches:          64
Percent Similarity:     53.29%          Conservative:     25
Best Local Similarity:  38.32%          Mismatches:       50
Query Match:            12.54%          Indels:           28
DB:                     4               Gaps:             3

US-10-644-659A-1 (1-1146) x ABB60049 (1-399)

Qy  718 AGCCCAGTGGCAACTTGAAAGGAGGATGGCAGCAGTGGGCTGATGAACACATACAATCC 777
            :|::::                  ::                :::|||  |:::
Db   22 SerProLeuSerSerLysValAlaMetPheAsnAsnGlnAlaThrGlnHisLysGlnSer 41

Qy  778 CAGAAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGGTCGCCATGTCCACCCGCCTA 837
        |||  || |||||||||||                      ||| | |   :
Db   42 GlnLeuLeuAsnProPheSerGlnAsp-------GlyArgAlaAlaSerProLysPro 58

Qy  838 CACAAAGGAGGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCC 897
            :: |||                |||   :||| :|||
Db   59 ThrPheSerLysAspGlnTyrGlyLysProLeuAlaGlySerLeuThrGlnMetArgGly 78

Qy  898 AAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACA 957
        ::::|||     || :::| :  |||     ||||||||   || :: |  :::
Db   79 GlnLysAlaAsnIleHisValMetLysGluMetLeuGluLeuCysGlnIleIleAsnSer 98

Qy  958 ATGGCTCGCCACCAGACGAGATGCAAGATCCAG------GTTACTTTGGAGATCTCTTT 1011
            :::||                    ||| |
Db   99 GluGlyTyrArgValLysAspGluProThrMetArgValIleProPheGlyLeuLeuPhe 118

Qy 1012 GAC--------------------------------------------------- 1014
            :::
Db  119 AsnValSerValLeuPheThrAlaGlyIlePhePheGluLysProSerLysValThr 138

Qy 1015 ---------AGATACGTTCGTATTTCAGATAAAGTAGTAGTGGGCCATTCTCATGCGTGCC 1062
```

Db   20 PheAsnGlnGlnAlaGluGlnHisLysAsnTrpMetMetIleAsnProPheAlaHis--- 38

Qy   805 TTTGATTACGAGCTGGCCATGTCCACCGCCTACCACAAAGGAGATGAGGGCTATGGCCGC 864
               |||| :::|| :: |||||||| :::||| ||||||||
Db   39 -----TyrAsnValAsnGluMetProLysArgThrPheProGluGluTyrGlyArg 56

Qy   865 CCCAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCTGCTGAGGAGCACATCTACAGG 924
               ||||::: ::: :::||||::: ::||
Db   57 AlaProThrGlySerLeuSerGluGlnArgSerLeuGlnAlaAsnValArgAlaLeuGlu 76

Qy   925 GAAATGATGGACATGTGCTTCATTATCTGCACAATGGCTGCCACAGACGA---GATGGC 981
               |||::: || :::||| || ||| |
Db   77 GluIleLeuGlnLeuCysLysAspLeuIleGlnLysSerGlyArgAspAspProIleAspGly 96

Qy   982 AAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATACGTTCGTATTCAGATAAAGTA 1041
               ::: ||||||| ||||||::: ||||| ||| ::|||::::
Db   97 ArgLysValLeuAlaPheGlyGlnLeuPheGluThrTyrAsnAsnIleSerAspLysLeu 116

Qy   1042 GTGGGCATTCTCATGCGTGCCAGGAAACATGGACTTGGTAGACTTTGAAGGAGAGATGCTA 1101
               ::: |||||||::: | ||| :::||| ||
Db   117 LeuAlaThrLeuLeuGlyAlaArgLysTyrGlyPheValAspPheSerGlyGluThrLeu 136

Qy   1102 TGGCAAGGCCGAGATGACCATCGTTGTGATTACGCTACTCAAG 1143
               :::||||||| :::|||| |
Db   137 PheGlnGlyArgAspAspThrGluProValArgLeuLeuArg 150

RESULT 14
AAM21625
ID   AAM21625 standard; protein; 37 AA.
XX
AC   AAM21625;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Peptide #8059 encoded by probe for measuring cervical gene expression.
XX
KW   probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-0063366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 27; SEQ ID NO 26451; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably

---

Db   139 SerThrLeuGlnIleTyrAsnTyrIleSerArgAspLysValValGlyIleLeuLeuArgAla 158
               || |||||||||||||||||||||
Qy   1063 AGGAAACATGGACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGATGACCAT 1122
               |||||||| ||||||||||||||:::||| ||
Db   159 ArgLysHisLysValAspPheGluGlyGluMetLeuTrpGlnArgArgAspAspThrAsp 178

Qy   1123 GTTGTGATTACGCTACTCAAG 1143
               ::: ||||||:::|||||
Db   179 ValProValPheLeuLeuLys 185

RESULT 13
ABB58876
ID   ABB58876 standard; protein; 183 AA.
XX
AC   ABB58876;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3420.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02979.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 3420; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 183 AA;

Alignment Scores:
Pred. No.:   3.17e-10
Score: 205.00
Percent Similarity: 55.22%   Length: 183
Best Local Similarity: 35.07%   Matches: 47
Query Match: 9.96%   Conservative: 27
DB: 4   Mismatches: 56
                                                  Indels: 4
                                                  Gaps: 2

US-10-644-659a-1 (1-1146) x ABB58876 (1-183)

Qy   745 TGGCAGCAGTGGGCTGATGAACACATACCCCAGAAGCTCAATCCTTTCAGTGAAGAG 804
               ::: ||| |||||::|| :::|||||||::::

```
CC   cervical cancer. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 37 AA;

Alignment Scores:
Pred. No.:
Score:                         4.38e-09
                               190.00
Percent Similarity:            97.30%
Best Local Similarity:         97.30%
Query Match:                   9.23%
DB:                            4

US-10-644-659A-1 (1-1146) x AAM21625 (1-37)

Qy    260 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LysAspMetGluMetAspLysAlaGlnArgLysProLeuArgPheLeuThrSerLysArg 20

Qy    320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACG 370
          ||||||||||||||||||||||||||||||| ||||||||||||||||
Db     21 LysArgCysProLysArgTrpSerAlaArgLeuMetArgGluGluGlyThr 37

Length:        37
Matches:       36
Conservative:  0
Mismatches:    1
Indels:        0
Gaps:          0


RESULT 15
ABB43985
ID   ABB43985 standard; peptide; 37 AA.
XX
AC   ABB43985;
XX
DT   04-FEB-2002  (first entry)
XX
DE   Peptide #11491 encoded by human foetal liver single exon probe.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 27; SEQ ID NO 36620; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a peptide encoded by a single exon nucleic acid probe
CC   of the invention. Note: The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 37 AA;
```

```
Alignment Scores:
Pred. No.:
Score:                         4.38e-09
                               190.00
Percent Similarity:            97.30%
Best Local Similarity:         97.30%
Query Match:                   9.23%
DB:                            4

US-10-644-659A-1 (1-1146) x ABB43985 (1-37)

Qy    260 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LysAspMetGluMetAspLysAlaGlnArgLysProLeuArgPheLeuThrSerLysArg 20

Qy    320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACG 370
          ||||||||||||||||||||||||||||||| ||||||||||||||||
Db     21 LysArgCysProLysArgTrpSerAlaArgLeuMetArgGluGluGlyThr 37

Length:        37
Matches:       36
Conservative:  0
Mismatches:    1
Indels:        0
Gaps:          0


Search completed: May  4, 2004, 08:51:36
Job time : 70 secs
```

```
                    GenCore version 5.1.6

          Copyright (c) 1993 - 2004 Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:       May  5, 2004, 00:04:42 ; Search time 111 Seconds
                                  (without alignments)
                          5729.493 Million cell updates/sec

Title:           US-10-644-659A-1
Perfect score:   1146
Sequence:        1 atggctccgggcgaaagga..........tgattacgctactccagtga 1146

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:     1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
                 2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
                 3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
                 4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
                 5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
                 6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	72	6.3	7218	1	US-08-232-463-14	Sequence 14, Appl
2	39.2	3.4	289	3	US-09-007-005-17	Sequence 17, Appl
3	39.2	3.4	289	3	US-09-244-796-17	Sequence 17, Appl
4	38.6	3.4	5751	4	US-09-023-655-1415	Sequence 1415, Ap
5	38.6	3.4	11907	1	US-08-061-376-4	Sequence 4, Appli
6	38.6	3.4	14255	1	US-08-320-559-1	Sequence 1, Appli
7	38.6	3.4	14255	1	US-08-327-392-1	Sequence 1, Appli
8	38.6	3.4	14255	5	US-08-306-691B-55	Sequence 55, Appl
9	38.6	3.4	14255	3	US-08-545-860D-1	Sequence 1, Appli
10	38.6	3.4	14255	5	PCT-US94-04496-1	Sequence 1, Appli
11	38.6	3.2	729	4	US-09-540-236-357	Sequence 357, App
12	36.2	3.2	1136	1	US-08-395-800A-9	Sequence 9, Appli
13	36.2	3.2	1155	1	US-08-434-151-1	Sequence 1, Appli
14	36.2	3.2	1155	2	US-08-208-889A-1	Sequence 1, Appli
15	36.2	3.2	1155	2	US-08-433-27-1	Sequence 1, Appli
16	36.2	3.2	1155	2	US-08-715-259-1	Sequence 1, Appli
17	36.2	3.2	1174	5	PCT-US95-07554-3	Sequence 3, Appli
18	36.2	3.2	1199	3	US-08-395-800A-5	Sequence 5, Appli
19	36.2	3.2	2268	3	US-08-675-773B-4	Sequence 4, Appli
20	36.2	3.2	3373	1	US-08-273-411-2	Sequence 2, Appli
21	36.2	3.2	3791	3	US-08-675-773B-3	Sequence 3, Appli
22	36.2	3.2	8174	1	US-07-914-281-5	Sequence 5, Appli
23	36.2	3.2	8174	1	US-08-393-246-5	Sequence 5, Appli
24	36.2	3.2	8174	1	US-08-525-058A-5	Sequence 5, Appli
25	36.2	3.2	8174	5	US-08-696-731-5	Sequence 5, Appli
26	36.2	3.2	8174	3	US-09-042-531-5	Sequence 5, Appli
27	36.2	3.2	8174	5	PCT-US91-00899-3	Sequence 3, Appli
28		3.1			2781	4	US-09-342-681C-116	Sequence 116, App
29		3.1			4235	4	US-09-342-681C-18	Sequence 18, Appl
30		3.1			589	3	US-09-328-111-149	Sequence 149, App

No.		%	Length	DB	Score	ID	Description
28	c	36	3.1	2781	4	US-09-342-681C-116	Sequence 116, App
29	c	36	3.1	4235	4	US-09-342-681C-18	Sequence 18, Appl
30		35	3.0	589	3	US-09-328-111-149	Sequence 149, App
31	c	34.8	3.0	53526	3	US-08-658-136-2	Sequence 2, Appli
32		34.8	3.0	53577	3	US-08-658-136-1	Sequence 1, Appli
33		34.6	3.0	99629	3	US-09-596-002-37	Sequence 37, Appl
34		34.4	3.0	2832	3	US-08-819-177-2	Sequence 2, Appli
35	c	34.2	3.0	1864	1	US-08-353-948-1	Sequence 1, Appli
36		33.6	2.9	364	4	US-09-621-976-17202	Sequence 17202, A
37		33.4	2.9	1926	4	US-09-249-585A-2	Sequence 2, Appli
38		33.4	2.9	1926	3	US-09-410-399-3	Sequence 3, Appli
39		33.4	2.9	2580	4	US-09-050-863-2	Sequence 2, Appli
40		33.4	2.9	2580	4	US-09-355-081-2	Sequence 2, Appli
41		33.4	2.9	3914	1	US-08-117-373-11	Sequence 11, Appl
42	c	33.4	2.9	5452	2	US-09-130-114-1	Sequence 1, Appli
43	c	33.4	2.9	8705	4	US-09-647-344A-14	Sequence 14, Appl
44		33.4	2.9	9600	4	US-09-910-647-1	Sequence 1, Appli
45		33.4	2.9	9600	4	US-09-620-925-1	Sequence 1, Appli

```
                              ALIGNMENTS

RESULT 1
US-08-232-463-14/c
; Sequence 14, Application US/0823246 3
; Patent No. 5670367
;~ GENERAL INFORMATION:
;~ APPLICANT: DORNER, F.
;~ APPLICANT: SCHEIFLINGER, F.
;~ APPLICANT: FALKNER, F. G.
;~ TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;~ NUMBER OF SEQUENCES: 52
;~ CORRESPONDENCE ADDRESS:
;~ ADDRESSE: Foley & Lardner
;~ STREET: 1800 Diagonal Road, Suite 500
;~ CITY: Alexandria
;~ STATE: VA
;~ COUNTRY: USA
;~ ZIP: 22313-0299
;~ COMPUTER READABLE FORM:
;~ MEDIUM TYPE: Floppy disk
;~ COMPUTER: IBM PC Compatible
;~ OPERATING SYSTEM: PC-DOS/MS-DOS
;~ SOFTWARE: PatentIn Release #1.0, Version #1.25
;~ CURRENT APPLICATION DATA:
;~ APPLICATION NUMBER: US/08/232,463
;~ FILING DATE:
;~ CLASSIFICATION: 435
;~ PRIOR APPLICATION DATA:
;~ APPLICATION NUMBER: US/07/935,313
;~ FILING DATE:
;~ APPLICATION NUMBER: EP 91 114 300.6
;~ FILING DATE: 26-AUG-1991
;~ ATTORNEY/AGENT INFORMATION:
;~ NAME: BENT, Stephen A.
;~ REGISTRATION NUMBER: 29,768
;~ REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;~ TELECOMMUNICATION INFORMATION:
;~ TELEPHONE: (703)836-9300
;~ TELEFAX: (703)683-4109
;~ TELEX: 899149
;~ INFORMATION FOR SEQ ID NO: 14:
;~ SEQUENCE CHARACTERISTICS:
;~ LENGTH: 7218 base pairs
;~ TYPE: nucleic acid
;~ STRANDEDNESS: single
;~ TOPOLOGY: linear
;~ IMMEDIATE SOURCE:
;~ CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match             6.3%;    Score 72;  DB 1;  Length 7218;
```

```
Best Local Similarity  4.5%;  Pred. No. 4.3e-11;
Matches  18;  Conservative  237;  Mismatches 147;  Indels  0;  Gaps  0;

Qy   274 GGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAA 333
          :: :::  ::::: : :::::: : ::: ::RRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1438 GTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

Qy   334 ACGGTGGTCAGCAAGACTTACGAGAGAGGGACGTGAGCCACCTCAGCCACAGGTAC 393
          : :::::  :::::::::::::: :: : ::  :: :: ::::
Db  1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319

Qy   394 GAGAGGGATGCTGTGTGTCTTGAACCTGGCAGCCAGCAGAATGACATTGACAGAATCCTC 453
          :::::: :  :: ::: ::::: :: :: : : : : :: : : : :::::
Db  1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259

Qy   454 CACAGCCCACGCTCCCCAACGCGGAGAAAATGTGCCAACCTGGTCTCTGAGCTAACC 513
          :::::::::::::::::::::::RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
Db  1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199

Qy   514 AAGGGCTGAGAGTGATGGAGCAGGAGGCCCACATGGAGAGTGACAGCGTAGACACA 573
          ::::  : :::: : ::::::: : : : ::: : : ::  :::: : : ::::
Db  1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139

Qy   574 GAGACAGCGCGCTATGGAGGAGGCTGAGGAGCCCGAGCAGGATGGAGTGCAGGTG 633
          :::::: :: :: ::: :::: :: :: : : ::: : : :: : ::::
Db  1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079

Qy   634 GCTGTGGTCAGGATCAAGCGCCCCTTGCCCTCCCAGGTAAAC 675
          : :: :  ::: ::  ::::|||||  |||||  ||||||
Db  1078 RRRRRRRRRRRRRRRATCGCAAGCTCCCTCGACCTGCGAGCCAAGC 1037
```

RESULT 2
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258858
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```
Query Match   3.4%;  Score 39.2;  DB 3;  Length 289;
Best Local Similarity  3.7%;  Pred. No. 0.049;  Gaps  0;
Matches  6;  Conservative  82;  Mismatches  74;  Indels  0;  Gaps  0;

Qy   493 AACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAGGCCCACATGG 552
          ||| : :: ::|:| :|: : |: : : : : : : : :: : : : ::
Db    33 RARCRARARURURARCRARARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 92
Qy   553 AGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGGAGAGGCCC 612
          : : :: : : : :: :: : : : :: : : ::
```

```
Qy   613 GAGCAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGC 654
          : : : : :: : : :: : : :: : :: : : ::
Db   153 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 194
```

RESULT 3
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

```
Query Match   3.4%;  Score 39.2;  DB 3;  Length 289;
Best Local Similarity  3.7%;  Pred. No. 0.049;  Gaps  0;
Matches  6;  Conservative  82;  Mismatches  74;  Indels  0;  Gaps  0;

Qy   493 AACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAGGCCCACATGG 552
          ||| : :: ::|:| :|: : |: : : : : : : : :: : : : ::
Db    33 RARCRARARURURARCRARARURGRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 92
Qy   553 AGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGGAGAGGCCC 612
          : : :: : : : :: :: : : : :: : : ::
Db    93 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 152
Qy   613 GAGCAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGC 654
          : : : : :: : : :: : : :: : :: : : ::
Db   153 RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS 194
```

RESULT 4
US-09-023-655-1415/c
; Sequence 1415, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/061,376
FILING DATE: 13-MAY-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9387
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)546-4737
TELEFAX: (619)546-9392
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 11907 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-061-376-4

Query Match   3.4%;   Score 38.6;   DB 3;   Length 11907;
Best Local Similarity   51.4%;   Pred. No. 0.57;
Matches   89; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   465 CTCCCAACGCGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   405 CTCCCCAAACACGGCCCGGAACCGGCCCAGGTTGGTGCCGAGACCTGCAG 346

Qy   525 AGTGATGGAGCAGGAGGAGCCCACATGAGGAGTGACACGCGTAGACACAGAGGACAGCGG 584
          |||||||||||||||||||||||||||||||||||||||||||||
Db   345 CGCCGCGTCGAAGCCCGGCCCACCCGGAGCAGGCCCGCCCTGAAGAGGGCTGAGGACGA 286

Qy   585 CTATGGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTG 637
          ||||||||||||||||||||||||||||||||||||||
Db   285 TGACGAAGACGAAGACGAGGCGGACGACGAGAGGCGTGCTGAGGCGGCGGCCG 233

RESULT 6
US-08-320-559-1/c
; Sequence 1, Application US/08320559
; Patent No. 5633135
; GENERAL INFORMATION:
APPLICANT: Croce, Carlo
APPLICANT: Canaani, Eli
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
TITLE OF INVENTION: All-1 Region
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM PS/2
OPERATING SYSTEM: PC-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/320,559
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/062,443
FILING DATE: 14 MAY 1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/971,094
FILING DATE: 30-OCT-92
PRIOR APPLICATION DATA:

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,655
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1415:
SEQUENCE CHARACTERISTICS:
LENGTH: 5751 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g467511
US-09-023-655-1415

Query Match   3.4%;   Score 38.6;   DB 4;   Length 5751;
Best Local Similarity   51.4%;   Pred. No. 0.38;
Matches   89; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   465 CTCCCAACGCGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   428 CTCCCCAAACACGGCCCGGAACCGGCCCAGGTTGGTGCCGATGGCCGAGACCTGCAG 369

Qy   525 AGTGATGGAGCAGGAGGAGCCCACATGAGGAGTGACACGCGTAGACACAGAGGACAGCGG 584
          |||||||||||||||||||||||||||||||||||||||||||||
Db   368 CGCCGCGTCGAAGCCCGGCCCACCCGGAGCAGGCCCGCCCTGAAGAGGGCTGAGGACGA 309

Qy   585 CTATGGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTG 637
          ||||||||||||||||||||||||||||||||||||||
Db   308 TGACGAAGACGAAGACGAGGCGGACGACGAGAGGCGTGCTGAGGCGGCGGCCG 256

RESULT 5
US-08-061-376-4/c
; Sequence 4, Application US/08061376
; Patent No. 6175000
; GENERAL INFORMATION:
APPLICANT: Evans, Glen A.
APPLICANT: Djabali, Malek
APPLICANT: Selleri, Licia
APPLICANT: Parry, Pauline
TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23
TITLE OF INVENTION: TRANSLOCATION BREAKPOINT ASSOCIATED WITH ACUTE LEUKEMIAS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```
          APPLICATION NUMBER: US/07/888,830
          FILING DATE: 27-MAY-92
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: US/07/805,093
          FILING DATE: 11-DEC-91
     ATTORNEY/AGENT INFORMATION:
          NAME: DeLuca, Mark
          REGISTRATION NUMBER: 33,229
          REFERENCE/DOCKET NUMBER: TJU-0855
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: (215) 568-3100
          TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
     SEQUENCE CHARACTERISTICS:
          LENGTH: 14255
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
          ANTI-SENSE: No
US-08-320-559-1

Query Match        3.4%;   Score 38.6;  DB 1;  Length 14255;
Best Local Similarity 51.4%;  Pred. No. 0.63;  Indels 0;  Gaps 0;
Matches  89;  Conservative 0;  Mismatches 84;

Qy   465 CTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524
         ||||||||   ||||  ||   ||  ||||  || |||    ||| |   ||| ||| ||
Db   219 CTCCCCAAACACGGCCCGGAACCGGCCCAGGTTGGTGCCGATGGCGGCCGAGACCTGCAG 160

Qy   525 AGTGATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACAGAGGACAGCGG   584
         ||  | || | |||  |  |||  |  || ||||| |  |  |    |||| | | |||
Db   159 CGCCGGTGCGAAGCCCGGGCCCACCCGGAGCAGGGCCCGGCCCTGAAGAGGCTGAGGACGA 100

Qy   585 CTATGGAGGAGAGGCTGAGGAGAGAGCCCGAGCAGGATGAGTGGAGGTCAGGTGGCTG   637
         | | || || || | |||| | || |||| || || |   | | ||||| | || ||
Db    99 TGACGAAGACGAAGACGAGGCGGACGACCAGGAGGCGGCTGCTGAGGCGGCGGCCG     47


RESULT 7
US-08-327-392-1/c
; Sequence 1, Application US/08327392
; Patent No. 5633136
GENERAL INFORMATION:
     APPLICANT: Croce, Carlo
     APPLICANT: Canaani, Eli
     TITLE OF INVENTION: ALL-1 Polynucleotides and Monoclonal
     TITLE OF INVENTION: Antibodies for Leukemia Detection and
     TITLE OF INVENTION: Treatment
     NUMBER OF SEQUENCES: 24
     CORRESPONDENCE ADDRESS:
          ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 563136ris
          STREET: One Liberty Place - 46th Floor
          CITY: Philadelphia
          STATE: PA
          COUNTRY: USA
          ZIP: 19103
     COMPUTER READABLE FORM:
          MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
          COMPUTER: IBM
          OPERATING SYSTEM: PC-DOS
          SOFTWARE: WORDPERFECT 5.1
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/327,392
          FILING DATE:
          CLASSIFICATION: 435
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 07/971,094
          FILING DATE: 30-OCT-92
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 07/888,830
          FILING DATE: 27-MAY-92
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER: 07/805,093
```

```
          FILING DATE: 11-DEC-91
     ATTORNEY/AGENT INFORMATION:
          NAME: DeLuca, Mark
          REGISTRATION NUMBER: 33,229
          REFERENCE/DOCKET NUMBER: TJU-1331
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: (215) 568-3100
          TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 1:
     SEQUENCE CHARACTERISTICS:
          LENGTH: 14255
          TYPE: nucleic acid
          STRANDEDNESS: single
          TOPOLOGY: linear
          ANTI-SENSE: No
US-08-327-392-1

Query Match        3.4%;   Score 38.6;  DB 1;  Length 14255;
Best Local Similarity 51.4%;  Pred. No. 0.63;  Indels 0;  Gaps 0;
Matches  89;  Conservative 0;  Mismatches 84;

Qy   465 CTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524
         ||||||||   ||||  ||   ||  ||||  || |||    ||| |   ||| ||| ||
Db   219 CTCCCCAAACACGGCCCGGAACCGGCCCAGGTTGGTGCCGATGGCGGCCGAGACCTGCAG 160

Qy   525 AGTGATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACAGAGGACAGCGG   584
         ||  | || | |||  |  |||  |  || ||||| |  |  |    |||| | | |||
Db   159 CGCCGGTGCGAAGCCCGGGCCCACCCGGAGCAGGGCCCGGCCCTGAAGAGGCTGAGGACGA 100

Qy   585 CTATGGAGGAGAGGCTGAGGAGAGAGCCCGAGCAGGATGAGTGGAGGTCAGGTGGCTG   637
         | | || || || | |||| | || |||| || || |   | | ||||| | || ||
Db    99 TGACGAAGACGAAGACGAGGCGGACGACCAGGAGGCGGCTGCTGAGGCGGCGGCCG     47


RESULT 8
US-08-306-691B-55/c
; Sequence 55, Application US/08306691B
; Patent No. 5734039
GENERAL INFORMATION:
     APPLICANT: Calabretta, Bruno
     APPLICANT: Skorski, Tomasz
     TITLE OF INVENTION: ANTISENSE
     TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
     NUMBER OF SEQUENCES: 55
     CORRESPONDENCE ADDRESS:
          ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.
          STREET: Two Penn Center, Suite 1800
          CITY: Philadelphia
          STATE: Pennsylvania
          COUNTRY: U.S.A.
          ZIP: 19102
     COMPUTER READABLE FORM:
          MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
          COMPUTER: IBM PS/2
          OPERATING SYSTEM: MS-DOS
          SOFTWARE: WordPerfect 5.1
     CURRENT APPLICATION DATA:
          APPLICATION NUMBER: US/08/306,691B
          FILING DATE: September 15, 1994
          CLASSIFICATION: 514
     PRIOR APPLICATION DATA:
          APPLICATION NUMBER:
          FILING DATE:
     ATTORNEY/AGENT INFORMATION:
          NAME: Monaco, Daniel A.
          REGISTRATION NUMBER: 30,480
          REFERENCE/DOCKET NUMBER: 8321-8
     TELECOMMUNICATION INFORMATION:
          TELEPHONE: (215) 568-8383
          TELEFAX: (215) 568-5549
          TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 55:
     SEQUENCE CHARACTERISTICS:
```

```
;     LENGTH: 14255 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
US-08-306-691B-55

Query Match             3.4%;   Score 38.6;   DB 1;   Length 14255;
Best Local Similarity  51.4%;   Pred. No. 0.63;
Matches   89;  Conservative   0;  Mismatches   84;  Indels   0;  Gaps   0;

Qy    465 CTCCCAACGCGGAGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTGGAG 524
          ||||||     |||    |||    ||       |||| |||||    |  |||| ||||||
Db    219 CTCCCAAACACGGCCCGGAACCGGCCCGAGGTTGGTGCCGATGGCGGCCGAGACCTGCAG 160

Qy    525 AGTGATGGAGCAGGAGGAGCCCACATGAGGAGTGACAGCGTAGACACAGAGGACAGCGG 584
          |  ||| ||||  || || || ||    ||||      |||  |||| ||||| || ||
Db    159 CGCCGGTCGAAGCCCGGGCCCACCCGGAGCAGGGCCCGAGCAGGATGGAGGTGAGGACGA 100

Qy    585 CTATGGAGGAGAGGCTGAGGAGGAGGCCCCGAGCAGGATGGAGTGCAGGTGGCTG 637
          |      |||  ||| ||||||    || |||      ||| |||| |  ||||
Db     99 TGACGAAGACGAAGACAGGCCGGACGACGACGAGGCGGCTGCTGAGGCGGCGGCCG 47

RESULT 9
US-08-545-860D-1/c
; Sequence 1, Application US/08545860D
; Patent No. 6040140
; GENERAL INFORMATION:
;     APPLICANT: Croce, Carlo
;     APPLICANT: Canaani, Eli
;     TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
;     TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
;     TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region
;     NUMBER OF SEQUENCES: 94
;     CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewitz &
;     ADDRESSEE: No. 6040140ris
;     STREET: One Liberty Place, 46th floor
;     CITY: Philadelphia
;     STATE: Pennsylvania
;     COUNTRY: USA
;     ZIP: 19103
;     COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/545,860D
;     FILING DATE: 07-MAR-1996
;     CLASSIFICATION: 435
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US94/04496
;     FILING DATE: 22-APR-1994
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US92/10930
;     FILING DATE: 09-DEC-1992
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/327,392
;     FILING DATE: 19-OCT-1994
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/320,559
;     FILING DATE: 11-OCT-1994
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/062,443
;     FILING DATE: 14-MAY-1993
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/971,094
;     FILING DATE: 30-OCT-1992
;     PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 07/888,839
;     FILING DATE: 27-MAY-1992
;     PRIOR APPLICATION DATA:
```

```
;     APPLICATION NUMBER: US 07/805,093
;     FILING DATE: 11-DEC-1991
;     ATTORNEY/AGENT INFORMATION:
;     NAME: DeLuca Esq., Mark
;     REGISTRATION NUMBER: 33,229
;     REFERENCE/DOCKET NUMBER: TJU-1262
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (215) 568-3100
;     TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 14255
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     ANTI-SENSE: No
US-08-545-860D-1

Query Match             3.4%;   Score 38.6;   DB 3;   Length 14255;
Best Local Similarity  51.4%;   Pred. No. 0.63;
Matches   89;  Conservative   0;  Mismatches   84;  Indels   0;  Gaps   0;

Qy    465 CTCCCAACGCGGAGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAGGGCTGGAG 524
          ||||||     |||    |||    ||       |||| |||||    |  |||| ||||||
Db    219 CTCCCAAACACGGCCCGGAACCGGCCCGAGGTTGGTGCCGATGGCGGCCGAGACCTGCAG 160

Qy    525 AGTGATGGAGCAGGAGGAGCCCACATGAGGAGTGACAGCGTAGACACAGAGGACAGCGG 584
          |  ||| ||||  || || || ||    ||||      |||  |||| ||||| || ||
Db    159 CGCCGGTCGAAGCCCGGGCCCACCCGGAGCAGGGCCCGAGCAGGATGGAGGTGAGGACGA 100

Qy    585 CTATGGAGGAGAGGCTGAGGAGGAGGCCCCGAGCAGGATGGAGTGCAGGTGGCTG 637
          |      |||  ||| ||||||    || |||      ||| |||| |  ||||
Db     99 TGACGAAGACGAAGACAGGCCGGACGACGACGAGGCGGCTGCTGAGGCGGCGGCCG 47

RESULT 10
PCT-US94-04496-1/c
; Sequence 1 Application PCT/US9404496
; GENERAL INFORMATION:
;     APPLICANT: Croce, Carlo
;     APPLICANT: Canaani, Eli
;     TITLE OF INVENTION: Diagnostics, Therapeutics and Methods
;     TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias
;     TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1
;     NUMBER OF SEQUENCES: 86
;     CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewitz &
;     ADDRESSEE: Norris
;     STREET: One Liberty Place, 46th floor
;     CITY: Philadelphia
;     STATE: Pennsylvania
;     COUNTRY: USA
;     ZIP: 19103
;     COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US94/04496
;     FILING DATE:
;     CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;     NAME: DeLuca Esq., Mark
;     REGISTRATION NUMBER: 33,229
;     REFERENCE/DOCKET NUMBER: TJU-1242
;     TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (215) 568-3100
;     TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 14255
;     TYPE: nucleic acid
```

```
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     ANTI-SENSE: No
PCT-US94-04496-1

Query Match        3.4%;   Score 38.6;  DB 5;  Length 14255;
Best Local Similarity 51.4%;   Pred. No. 0.63;  Indels 0;  Gaps 0;
Matches 89;   Conservative 0;  Mismatches 84;

Qy  465 CTCCCCAACGCGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524
        |||||||||  |  ||   ||  |||  || || |||| | |||   |     | |||
Db  219 CTCCCCAAACACGGCCCGGAACCGGCCCAGGTTGGTGCCGATGGCGCCGAGACCTGCAG 160

Qy  525 AGTGATGAGCAGAGAGGACGAGAGGAGTGACAGCGTAGACACAGAGGACAGCGG 584
        |  |    |||| || | || ||| || ||  |   | |||||   | ||| |
Db  159 CGCCGGTCGAAGCCCCGGGCCCACCCCGAGCAGAGGCGCCGGCCCTGAAGAGGCTGAGGACGA 100

Qy  585 CTATGGAGGAGGAGGCTGAGGAGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTG 637
        | |  ||  ||  || |  ||  ||  | |||| || |||| | |  ||  ||
Db  99  TGACGAAGAAGACGAGGACGAGGCGGCGACGACGAGGAGGCTGCTGAGGCGCGGCCG 47
```

RESULT 11
US-09-540-236-357/c
; Sequence 357, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; TITLE OF INVENTION: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 357
; LENGTH: 729
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-357

```
Query Match        3.2%;   Score 36.2;  DB 4;  Length 729;
Best Local Similarity 57.5%;   Pred. No. 0.64;  Indels 0;  Gaps 0;
Matches 65;   Conservative 0;  Mismatches 48;

Qy  425 AGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAA 484
        || |  | | | || | ||  | |  || ||   | ||| | || ||  |   |   |
Db  695 AGCACGGCCTATCCTCATTCATTCATCAAAAAATCCACATCCACGCCATCATAAGCAAAGCAAA 636

Qy  485 AATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAG 537
        | ||||||  || |  |||| |||  |  || |  |   |  |  |||| |||
Db  635 ATGGGTGCCCACCTCATCTTGACGGTAAGCAGGTCTTGGGTCTTGAGCGATAAG 583
```

RESULT 12
US-08-395-800A-9/c
; Sequence 9, Application US/08395800A
; Patent No. 5807732
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B
; APPLICANT: LENNON, GREGORY
; APPLICANT: ROUQUIER, SYLVIE
; APPLICANT: GIORGI, DOMINIQUE
; APPLICANT: KELLY, ROBERT J
; TITLE OF INVENTION: GDP-L-FUCOSE: BETA-D-GALACTOSIDE
; TITLE OF INVENTION: 2-ALPHA-L-FUCOSYLTRANSFERASES, DNA SEQUENCES ENCODING THE
; TITLE OF INVENTION: SAME, METHOD FOR PRODUCING THE SAME AND A METHOD OF
; TITLE OF INVENTION: GENOTYPING A PERSON
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA

; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/395,800A
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: 435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1136 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..1133
US-08-395-800A-9

```
Query Match        3.2%;   Score 36.2;  DB 1;  Length 1136;
Best Local Similarity 50.9%;   Pred. No. 0.82;  Indels 0;  Gaps 0;
Matches 86;   Conservative 0;  Mismatches 83;

Qy  472 ACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATG 531
        ||  |||  || | | |||| ||| | | |   ||     ||  ||  |   ||   |
Db  215 ACCCGGCCAGGCAGAGAGTGGCCACTGGGGGTGTCACCAGGCGGCGGTCTGGACACAGGAT 156

Qy  532 GAGCAGGAGGAGCCCACATGGAGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGA 591
        | |||||  |  |||  ||  || |  ||||  | | |   |||  |||||   |  ||||||
Db  155 CGACAGGGCCTAGGCCATGTGGAAAGTGTCTTGATCGATATGGAGGAAGAAGATTACAGA 96

Qy  592 GGAGAGGCTGAGGAGAGCCCGAGCAGGATGGAGTGCAGGTGGCTGTGG 640
        | |||  | || || | ||  || ||  | |||| ||| | | ||  |
Db  95  GACGAACACAGACTAGCAGGAGGCCAGGCAGAGCTGACGATGGCTCCGG 47
```

RESULT 13
US-08-434-151-1/c
; Sequence 1, Application US/08434151
; Patent No. 5700671
; GENERAL INFORMATION:
; APPLICANT: Prieto, P.
; APPLICANT: Smith, D.
; APPLICANT: Cummings, R.
; APPLICANT: Kopchick, J.
; APPLICANT: Mukerji, P.
; APPLICANT: Moremen, K.
; APPLICANT: Pierce, J.
; TITLE OF INVENTION: Transgenic Animals Producing Oligosaccharides and Glycocor
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Donald O. Nickey
; ADDRESSEE: ROSS Products Division
; ADDRESSEE: Abbott Laboratories
; STREET: 625 Cleveland Avenue
; CITY: Columbus
; STATE: Ohio
; COUNTRY: United States
; ZIP: 43215
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh System 7.1
; SOFTWARE: ClarisWorks 1.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434,151

```
;   FILING DATE: 02-MAY-1995
;   CLASSIFICATION: B00
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/209,132
;   FILING DATE: 09-MAR-1994
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (614) 624-7080
;   TELEFAX: (614) 624-3074
;   TELEX: No. 5700671e
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1155 base pairs
;     TYPE: Nucleic acid
;     STRANDEDNESS: Single
;     TOPOLOGY: Unknown
;   MOLECULE TYPE: Cloned cDNA representing the product of a human
;   MOLECULE TYPE: genomic DNA segment
;   DESCRIPTION: GDP-L-fucose-'-D-galactoside 2-alpha-fucosyl-    transferase
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   FRAGMENT TYPE: Entire amino acid sequence provided.
;   ORIGINAL SOURCE: Human Epidermal Carcinoma Cell line
;     ORGANISM:
;     STRAIN:
;     INDIVIDUAL ISOLATE:
;     DEVELOPMENTAL STAGE:
;     HAPLOTYPE:
;     TISSUE TYPE:
;     CELL TYPE:
;     CELL LINE:
;     ORGANELLE:
;   IMMEDIATE SOURCE: Human Epidermal Carcinoma Cell line
;     LIBRARY:
;     CLONE:
;   POSITION IN GENOME:
;     CHROMOSOME/SEGMENT: 19
;     MAP POSITION:
;     UNITS:
;   FEATURE:
;     NAME/KEY:
;     LOCATION:
;   IDENTIFICATION METHOD: DNA sequencing and restriction analysis
;   OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO: 1: is the enzyme
;   PUBLICATION INFORMATION:
;     AUTHORS: V.P. Rajan et al
;     TITLE: Journal of Biological Chemistry
;     JOURNAL:
;     VOLUME: 264
;     ISSUE:
;     PAGES: 11158 - 11167
;     DATE: 1989
;     DOCUMENT NUMBER:
;     FILING DATE:
;     PUBLICATION DATE:
;     RELEVANT RESIDUES IN SEQ ID NO:
US-08-434-151-1

Query Match       3.2%;  Score 36.2;  DB 1;  Length 1155;
Best Local Similarity  50.9%;  Pred. No. 0.83;
Matches  86;  Conservative  0;  Mismatches  83;  Indels  0;  Gaps  0;

Qy    472 ACGGGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTCATG 531
              ||| || |||||   ||| ||| |  ||||| | ||| |||  ||  ||| |||  ||  |
Db    214 ACCCGCAGGCAGGAGATGCCACTGGGGTGTCACCAGGCGCCGGGTCTGGACACAGGAT 155

Qy    532 GAGCAGGAGGAGGAGCCCACATGGAGGAGTGACACGGTAGACACAGAGGACAGCCGCCTATGGA 591
              |  |  || ||| || |||    |||  || | ||| | || ||| |  || ||
Db    154 CGACAGGCCTAGGCCATGTGGAAAGCTGTCTTGATGGATATGGAGGAAGAAGATTACACAGA 95

Qy    592 GGAGAGGCTGGAGGAGGAGCCCCGACCAGGATGGGAGTGCAGGTGGCTGTGG 640
              || |  |  ||| || ||  |||    || ||| | ||| ||| |||  |
Db     94 GAGGACAACAGACTAGCAGGAAGGCCAGGCGAGCAGCAGGCTGACGATGGCTCCGG 46
```

```
RESULT 14
US-08-208-889A-1/c
; Sequence 1, Application US/08208889A
; Patent No. 5750176
; GENERAL INFORMATION:
;   APPLICANT: Prieto, P., Smith, D., Cummings, R., Kopchick, J., Mukerji, P.,
;   APPLICANT: Moremen, K., Pierce, J.
;   TITLE OF INVENTION: Transgenic Production of Oligosaccharides and
;   TITLE OF INVENTION: Glycoconjugates
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Donald O. Nickey
;     ADDRESSEE: ROSS Products Division
;     ADDRESSEE: Abbott Laboratories
;     STREET: 625 Cleveland Avenue
;     CITY: Columbus
;     STATE: Ohio
;     COUNTRY: United States of America
;     ZIP: 43215
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: MS-DOS Version 6.21
;     SOFTWARE: WordPerfect Version 6.0a
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/208,889A
;     FILING DATE: 09-MAR-1994
;     CLASSIFICATION: 424
;   PRIOR APPLICATION DATA: No. 5750176 applicable
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (614) 624-7080
;     TELEFAX: (614) 624-3074
;     TELEX: No. 5750176e
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1155 base pairs
;     TYPE: Nucleic acid
;     STRANDEDNESS: Single
;     TOPOLOGY: Unknown
;   MOLECULE TYPE: Cloned cDNA representing the product of a
;   MOLECULE TYPE: human genomic DNA segment
;   DESCRIPTION: GDP-L-fucose- -D-galactoside 2-alpha-fucosyl-transferase
;   HYPOTHETICAL:
;   ANTI-SENSE:
;   FRAGMENT TYPE: Entire amino acid sequence provided.
;   ORIGINAL SOURCE: Human Epidermal Carcinoma Cell line
;     ORGANISM:
;     STRAIN:
;     INDIVIDUAL ISOLATE:
;     DEVELOPMENTAL STAGE:
;     HAPLOTYPE:
;     TISSUE TYPE:
;     CELL TYPE:
;     CELL LINE:
;     ORGANELLE:
;   IMMEDIATE SOURCE: Human Epidermal Carcinoma Cell line
;     LIBRARY:
;     CLONE:
;   POSITION IN GENOME:
;     CHROMOSOME/SEGMENT: 19
;     MAP POSITION:
;     UNITS:
;   FEATURE:
;     NAME/KEY:
;     LOCATION:
;   IDENTIFICATION METHOD: DNA sequencing and restriction analysis
;   OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO:
;   OTHER INFORMATION: 1: is the enzyme, GDP-L-fucose- -D-galactoside 2-alpha-fuc
;   OTHER INFORMATION: which has the amino acid sequence described in SEQ. ID NO:
;   OTHER INFORMATION: is responsible for the synthesis of 2'-fucosyllactose.
;   PUBLICATION INFORMATION:
;     AUTHORS: V.P. Rajan, et al.
```

; TITLE:
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 264
; ISSUE:
; PAGES: 11158 - 11167
; DATE: 1989
; DOCUMENT NUMBER:
; FILING DATE:
; PUBLICATION DATE:
; RELEVANT RESIDUES IN SEQ ID NO:
US-08-208-889A-1

Query Match   3.2%; Score 36.2; DB 1; Length 1155;
Best Local Similarity   50.9%; Pred. No. 0.83;
Matches   86; Conservative   0; Mismatches   83; Indels   0; Gaps   0;

Qy   472 ACGCGGAGGAGAAAAATGTGCCAACCTGGTCTCGAGCTAACCAAGGGCTGGAGAGTGATG 531
          |||||| |||||| ||| ||||| ||||||||| || ||||||||||||| ||||
Db   214 ACCCGGCAGGCAGAGATGGCCACTGGGGGTGTCACCAGGCGGCGGGTCTGGACACAGGAT 155

Qy   532 GAGCAGGAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGA 591
          ||||| ||||| || ||| ||||| ||||| ||||| |||||||| || ||||||
Db   154 CGACAGGCCTAGGCCATGTGGAAAGCTGTCTTGATGGATATGGATGGAGGAAGATTACAGA 95

Qy   592 GGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGTGGCTGTGG 640
          |||||| ||||||| |||||| ||||||||||||||| |||||
Db   94 GAGGACAGACTAGCAGGAAGGCCCAGGCCAGCGAGAGCTGACGATGGCTCCGG 46

RESULT 15
US-08-433-271-1/c
; Sequence 1, Application US/08433271
; Patent No. 5891698
; GENERAL INFORMATION:
; APPLICANT: Prieto, P., Smith, D., Cummings, R., Kopchick, J., Mukerji, P.,
; APPLICANT: Moremen, K., Pierce, J.
; TITLE OF INVENTION: Humanized Milk
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Donald O. Nickey
;   ADDRESSEE: ROSS Products Division
;   ADDRESSEE: Abbott Laboratories
;   STREET: 625 Cleveland Avenue
;   CITY: Columbus
;   STATE: Ohio
;   COUNTRY: United States of America
;   ZIP: 43215
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: MS-DOS Version 6.21
;   SOFTWARE: WordPerfect Version 6.0a
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/433,271
;   FILING DATE: 02-MAY-1995
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/209,122
;   FILING DATE: 09-MAR-1994
;   CLASSIFICATION:
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (614) 624-7080
;   TELEFAX: (614) 624-3074
;   TELEX: NO. 5891698e
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1155 base pairs
;   TYPE: Nucleic acid
;   STRANDEDNESS: Single
;   TOPOLOGY: Unknown
;   MOLECULE TYPE: Cloned cDNA representing the product of a
;   MOLECULE TYPE: human genomic DNA segment
;   DESCRIPTION: GDP-L-fucose- -D-galactoside 2-alpha-fucosyl-transferase

; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE: Entire amino acid sequence provided.
; ORIGINAL SOURCE: Human Epidermal Carcinoma Cell line
;   ORGANISM:
;   STRAIN:
;   INDIVIDUAL ISOLATE:
;   DEVELOPMENTAL STAGE:
;   HAPLOTYPE:
;   TISSUE TYPE:
;   CELL TYPE:
;   CELL LINE:
;   ORGANELLE:
; IMMEDIATE SOURCE: Human Epidermal Carcinoma Cell line
;   LIBRARY:
;   CLONE:
; POSITION IN GENOME:
;   CHROMOSOME/SEGMENT: 19
;   MAP POSITION:
;   UNITS:
; FEATURE:
;   NAME/KEY:
;   LOCATION:
;   IDENTIFICATION METHOD: DNA sequencing and restriction analysis
;   OTHER INFORMATION: The encoded product of nucleotide SEQ ID NO:
;   OTHER INFORMATION: 1 is the enzyme, GDP-L-fucose- -D-galactoside 2-alpha-fuc
;   OTHER INFORMATION: which has the amino acid sequence described in SEQ. ID NO.
;   OTHER INFORMATION: is responsible for the synthesis of 2'-fucosyllactose.
; PUBLICATION INFORMATION:
;   AUTHORS: V.P. Rajan, et al.
;   TITLE:
;   JOURNAL: Journal of Biological Chemistry
;   VOLUME: 264
;   ISSUE:
;   PAGES: 11158 - 11167
;   DATE: 1989
;   DOCUMENT NUMBER:
;   FILING DATE:
;   PUBLICATION DATE:
;   RELEVANT RESIDUES IN SEQ ID NO:
US-08-433-271-1

Query Match   3.2%; Score 36.2; DB 2; Length 1155;
Best Local Similarity   50.9%; Pred. No. 0.83;
Matches   86; Conservative   0; Mismatches   83; Indels   0; Gaps   0;

Qy   472 ACGCGGAGGAGAAAAATGTGCCAACCTGGTCTCGAGCTAACCAAGGGCTGGAGAGTGATG 531
          |||||| |||||| ||| ||||| ||||||||| || ||||||||||||| ||||
Db   214 ACCCGGCAGGCAGAGATGGCCACTGGGGGTGTCACCAGGCGGCGGGTCTGGACACAGGAT 155

Qy   532 GAGCAGGAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTATGGA 591
          ||||| ||||| || ||| ||||| ||||| ||||| |||||||| || ||||||
Db   154 CGACAGGCCTAGGCCATGTGGAAAGCTGTCTTGATGGATATGGATGGAGGAAGATTACAGA 95

Qy   592 GGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGTGGCTGTGG 640
          |||||| ||||||| |||||| ||||||||||||||| |||||
Db   94 GAGGACAGACTAGCAGGAAGGCCCAGGCCAGCGAGAGCTGACGATGGCTCCGG 46

Search completed: May 5, 2004, 03:05:10
Job time : 115 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:    May  4, 2004, 08:49:08 ; Search time 24 Seconds
                                    (without alignments)
                                    9186.291 Million cell updates/sec

Title:          US-10-644-659A-1
Perfect score:  2058
Sequence:       1 atggctccggcgaaagga.........tgattacgctactcaagtga 1146

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       283366 seqs, 9191526 residues

Total number of hits satisfying chosen parameters:      566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10644659/runat_04052004_084903_28540/app_query.fasta_1.1287
-DB=PIR_78 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITs=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CGN_1_1_38 @runat_04052004_084903_28540 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   PIR_78:*
             1: pir1:*
             2: pir2:*
             3: pir3:*
             4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	308.5	15.0	162	2	T21860	hypothetical prote
2	228.5	11.1	289	2	T24421	hypothetical prote
3	133.5	6.5	161	2	T04981	hypothetical prote
4	133.5	6.5	773	2	F90537	lipoprotein [impor
5	124	6.0	803	2	JC7862	eukaryotic initiat
6	123	6.0	838	1	EEWTHW	glutenin, high mol
7	122	5.9	887	2	S70642	ubiquitin ligase N
8	121	5.9	1182	2	I48378	hairless protein -
9	120.5	5.9	848	2	S02262	glutenin high mole
10	120.5	5.9	3938	2	T42761	Bassoon protein -
11	119.5	5.8	458	2	S24457	hypothetical prote
12	118.5	5.8	815	2	B30843	glutenin high mole
13	117.5	5.7	830	2	S15720	glutenin high mole
14	116.5	5.7	1001	2	S74544	translation initia

15	116	5.6	992	2	T46337	hypothetical prote
16	115.5	5.6	753	2	JC2099	glutenin, high mol
17	115.5	5.6	789	2	A30843	glutenin high mole
18	115.5	5.6	3942	2	T42730	Bassoon protein -
19	115	5.6	646	2	S15901	chromogranin B pre
20	114.5	5.4	487	c	F72126	ct339 hypothetical
21	114.5	5.4	487	c	A81545	conserved hypothet
22	114.5	5.4	487	c	B86495	CT339 hypothetical
23	114.5	5.6	549	2	C87719	protein R19.6 [lim
24	114.5	5.5	791	2	JN0690	glutenin, high-mol
25	114	5.5	1274	2	JN0015	trp protein - frui
26	114	5.5	1275	2	JU0092	trp protein - frui
27	113.5	5.5	728	2	S26427	70K structural pro
28	113.5	5.5	853	2	T10591	hypothetical prote
29	113.5	5.5	1293	2	S42402	xeroderma pigmento
30	113	5.5	529	2	T45134	hypothetical prote
31	112.5	5.5	1356	2	T16754	hypothetical prote
32	112.5	5.5	1876	2	E97944	zinc metalloprotei
33	112.5	5.5	6642	2	T29757	protein UNC-89 - C
34	112	5.4	729	2	S68191	triadin - human
35	111.5	5.4	754	2	AC2807	OmpA family protei
36	111.5	5.4	754	2	B97586	hypothetical prote
37	111	5.4	677	1	CNHUB	chromogranin B pre
38	111	5.4	815	2	JN0689	glutenin, high-mol
39	111	5.4	1840	2	T29091	transitin - chicke
40	110.5	5.4	255	1	WUHU4B	homeotic protein H
41	110.5	5.4	1217	2	T42625	AF-4 protein - mou
42	110.5	5.4	2774	2	A43359	microtubule-associ
43	110	5.3	278	2	B39066	proline-rich prote
44	110	5.3	476	2	S57963	methyl CpG binding
45	110	5.3	1173	2	T42719	TPR-containing/SH2

ALIGNMENTS

```
RESULT 1
T21860
hypothetical protein F36F2.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T21860
R;Cottage, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19479
A;Accession: T21860
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-162 <WIL>
A;Cross-references: EMBL:Z81532; PIDN:CAB04325.1; GSPDB:GN00019; CESP:F36F2.1
A;Experimental source: clone F36F2
C;Genetics:
A;Gene: CESP:F36F2.1
A;Map position: 1
A;Introns: 9/1; 28/1; 64/1; 108/3

Alignment Scores:
Pred. No.:                              1.78e-15          Length:         162
Score:                                  308.50            Matches:        61
Percent Similarity:                     70.54%            Conservative:   30
Best Local Similarity:                  47.29%            Mismatches:     31
Query Match:                            14.99%            Indels:         7
DB:                                     2                 Gaps:           3

US-10-644-659A-1 (1-1146) x T21860 (1-162)

Qy  760 GATGAACAACATACAATCCCAG---AAGCTCAATCCTTCAGTGAAGAGTTTGATTACGAG 816
            :::::::::::::      :::||   ||| :::||::|||::|||:|::||:   ||:
Db   19 GluGlnAsnValAlaThrGlnSerLysAspAlaThrTyrSerLysAspThrGlnLys  38

Qy  817 CTGGCCATGTCACCCGCCTACACAAAGGAGATGAGGGCTATGGCCGCCCAAAGAAGGA 876
            ||||||||||||||   ||:||||||||||||:|||||:|||:|||||||||||||
Db   39 ---------------LysMetAspLysSerSerGluTyrGlyArgProLysProGly  53
```

```
Qy   877  ACCAAAACTGCTGAAGGGCCAAGCGTGCTGAGGAGGAGCACATCTACAGGGAAAATGATGGAC  936
          ||||||||  ::|||||||||||||||  ||||||| :||
Db    54  ThrLeuThrGluGlnArgAlaLysLysAlaAlaAlaHisValHisArgGluMetLeuThr     73

Qy   937  ATGTGCTTCATTATCTGCACAATGGCTGCCCACAGACGAGATGGC---AAGATCCAGGTT     993
          ::: ||||    :::::: ||||||||||||||||
Db    74  LeuCysGluValValGluAspTyrGlyLysGlnGluLysGluGlyAspProIleArgIle    93

Qy   994  ACTTTTGGAGATCTCTTTGACAGATACGTTCGTATTTCGTATTTCAGATAAAGTAGTGGGCATTCTC  1053
          ||||||||                          :::
Db    94  ThrPheGlyArgLeuPheThrIleTyrValAsnIleSerAspLysValValGlyThrLeu  113

Qy  1054  ATGCGTGCCAGGAAACATGGACTGGTAGACTTTGAGGAGGAGATGCTATGGCAAGGCCGA  1113
          :::||||||  ||||||||||||| ::: |||
Db   114  LeuArgAlaArgLysHisGlyLeuValAspPheGluGluGluMetLeuPheGlnLysArg  133

Qy  1114  GATGACCAATGTTGTGATTACGTTACTC  1140
          ||||||   ||:::|||
Db   134  AspAspHisValIleIleThrLeuLeu  142
```

RESULT 2
T24421
hypothetical protein T04A8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24421
R;Palmer, S.
submitted to the EMBL Data Library, August 1994
A;Reference number: Z19889
A;Accession: T24421
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-289 <WIL>
A;Cross-references: EMBL:Z35563; PIDN:CAA84722.1; GSPDB:GN00021; CESP:T04A8
A;Experimental source: clone T04A8
C;Genetics:
A;Gene: CESP:T04A8.4
A;Map position: 3
A;Introns: 26/2; 58/3; 102/3; 191/2; 223/3

Alignment Scores:
Pred. No.:
Score: 2.07e-09   Length: 289
Percent Similarity: 228.50   Matches: 56
Best Local Similarity: 58.04%   Conservative: 27
Query Match: 39.16%   Mismatches: 49
DB: 11.10%   Indels: 11
                          2   Gaps: 4

US-10-644-659A-1 (1-1146) x T24421 (1-289)

```
Qy   721  CCAGTGGCAACTTGAAAGGG--------AGATGGCAGCAGTGGGCTGGGCTGATGAACACATA  771
          ||||||||::                  :::::
Db     2  ProIleGlySerAlaSerAspThrIleArgLysPheAsnAlaValAlaGlnAlaAsnGlu   21

Qy   772  CAATCCCAGAAGCTCAATCCTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACC  831
          ||||||                    ::: :::
Db    22  GluValLeuLysLysAsnProTyrSerAspThrTyrLysIleGln---AlaPheAspThr  40

Qy   832  CGCCTACACAAAGGAGAGTGAGGCTATGGCCGCCCCAAAGAGGAACCAAAACTGCTGAA  891
          ||||||||||||
Db    41  Lys----------------AsnTyrGlyArgProProGlySerLysThrGluAla  54

Qy   892  AGGGCCAAGCGTGCTGAGGAGGAGCACATCTACAGGGAAAATGATGGACATGTGCTTCATTATC  951
          |||    ::|||||||||
Db    55  ArgGlyIleLysAlaGlyValHisValCysArgGlyIleLeuPheLeuCysGluThrIle  74

Qy   952  TGCACAATGGCTGCCCACAGACGAGATGGCCAGATCCAGGTTACTTTTGGAGATCTCTTT  1011
          |||||||||||
Db    75  AspSerAsnAlaAspGlyGluGluProHisLys---TyrValLysPheGlyLysLeuPhe  93

Qy  1012  GACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACAT  1071
          :::  |||
Db    94  AsnIleTyrSerPheTyrSerAspLysSerValValGlyMetLeuValGlyLeuIleArgGluAlaArgLysTyr  113
```

RESULT 3
T04981
hypothetical protein T16L1.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C;Accession: T04981
R;Bevan, M.; Obermaier, B.; Deutschenbaur, S.; Piravandi, E.; Hoheisel, J.; Mewes, H.
submitted to the Protein Sequence Database, November 1998
A;Reference number: Z15393
A;Accession: T04981
A;Molecule type: DNA
A;Residues: 1-161 <BEV>
A;Cross-references: EMBL:AL031394
A;Experimental source: cultivar Columbia; BAC clone T16L1
C;Genetics:
A;Map position: 4
A;Introns: 93/3
A;Note: T16L1.130
C;Superfamily: Arabidopsis thaliana hypothetical protein T16L1.130

Alignment Scores:
Pred. No.:   Length: 161
Score: 0.0287   Matches: 35
Percent Similarity: 133.50   Conservative: 13
Best Local Similarity: 51.06%   Mismatches: 33
Query Match: 37.23%   Indels: 13
DB: 6.49%   Gaps: 2

US-10-644-659A-1 (1-1146) x T04981 (1-161)

```
Qy   862  CGCCCCAAAGAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGGAGCACATCTAC  921
          :::  |||||                      :::::  ||||||||||||
Db    62  LysGluGluArgMetAsnValAspGlyMetValAspGlyIleGlnLysGluGlnLysLeuGlnLysLeuGluLeuHis  81

Qy   922  AGGGAAAATGATGGACATGTGCTTCATTATCTGCACAATGGCTGCCCACAGACGAGATGGC  981
          |||                                        :::  |||||||
Db    82  Arg------------------------------------------LeuGlySerArgGlnThrAspGly  90

Qy   982  AAGATCCAGGTTACTTTTGGAGATCTCTTT------GACAGATACGTTCGTATTTCAGAT  1035
          :::||||||||||||||||||             |||||||||||||||||
Db    91  SerTyrLysValThrPheGlyValLeuPheAsnAspAspArgCysAlaAsnIlePheGlu  110

Qy  1036  AAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTTTGAGGAGGAG  1095
          :::|||                              :::|||
Db   111  AlaLeuValGlyThrLeuArgAlaAlaLysLysArgIleValAlaPheGluGluGlyGlu  130

Qy  1096  ATGCTATGGCAAGGCCGAGATGACCAATGTTGTGATTACGCTA  1137
          :::|||   |||||  |||||||||||
Db   131  LeuLeuLeuGlnGlyValHisAspLysValGlyValGluThrLeu  144
```

RESULT 4
F90537
lipoprotein [imported] - Mycoplasma pulmonis (strain UAB CTIP)
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: F90537
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: F90537
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-773 <KUR>

```
Qy  1072  GGACTGGTAGACTTTGAGGAGGAGATGCTATGGCAAGGCCGAGATGACCAATGTTGTGATT  1131
          ||||||||                                    ::|||| |||
Db   114  GlyLeuValHisPheGluGlyGluMetLeuTyrGlnArgGlnAspAspGluLysIleIle  133

Qy  1132  ACGTACTC  1140
          ||:::|||
Db   134  ThrMetLeu  136
```

A;Cross-references: GB:AL445566; PID:g14089619; PIDN:CAC13379.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU 2060
A;Genetic code: SGC3

Alignment Scores:
Pred. No.:   0.034   Length: 773
Score: 133.50   Matches: 80
Percent Similarity: 34.42%   Conservative: 47
Best Local Similarity: 21.68%   Mismatches: 130
Query Match: 6.49%   Indels: 113
                                 Gaps: 15

US-10-644-659A-1 (1-1146) x F90537 (1-773)

```
Qy   112 AATGAGAACACGCATCAGGCAGGCCAGGAGGCCTACAGGCTGGCTGCCGGGAGGGAGGACCCAG 171
             |||||  ||| ||||||    ::| ||||  ||||  |||||
Db    60 AsnArgAsnSerIleIleSerProGlnAsnPro--------------------------- 70

Qy   172 GACTCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCACACAGAAAGTCAGAGT 231
             |||||  :: |||:|| ||||||||||    ||||||||
Db    71 AspSerSerLysThrProGlnValProProThrLysProGluAspGlnAsnLys 90

Qy   232 GCCCCAAAGTCGCCACCCCGCCTGCCAGAAGGACATGGAAGGATGGACAAAGCTCAGAG--- 288
             ||| ||||||  :::|||||||      ::|||||
Db    91 GluProGlnSer----ProLysAspProGluIleLysAspAsnGlyGlnLysAsnGluGly 109

Qy   289 ---AAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAAGAGAGTGTC--------------- 329
            |||||||:::    ::: :|||||||||:|:
Db   110 SerLysAlaProGluIleLysAspMetSerGlnLysAspGlnAlaProGlnValProGln 129

Qy   329 -------------------------------------------------------------- 329
Db   130 LysGlnProGluAspProLysProGluThrGlnLnLysProProValLysSerGluAsp 149

Qy   330 CAAAACGGTGGTGTCAGCAAGACTTACGA------GAGAGGAGGGGACTGGAGCCACCTCAG 383
             ||||  |:||::::|:|||
Db   150 GlnAsnLysGluProGlnAspProLysGlnAlaProGluLysProLysAspProLysAsp 169

Qy   384 CCACAGTACCAGAGAGGATGCGTGTGCTTGAACCTGGGCAGCCAGA-------------- 431
            ||:||||:::
Db   170 ProGlnValLysAsp---------------MetAlaAsnLysAsnIleGlnGlyGly 182

Qy   432 --------------GAATGACATTGACAGAATCCTCCAGCCACGGCTCCCCAACGCGGAG 479
                           ||
Db   183 ProLysValProGluLysSerGlnAlaGlnThrProLysAspProGluIleLysAsnMetAsp 202

Qy   480 GAGAAA----------------------ATGTGCCAACCTGGTGTCTGAGCTAACCAAGGG 518
            :::|||                         |||||
Db   203 GlnLysAspGlnAlaProGlnLysProGlnLysGlnPro-----------GluAla 217

Qy   519 CTGGAGAGTGATGGAGCAGGAGGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGA 578
            ||||  ::||||||      :::|||
Db   218 SerLysLysPro-GluThrGlnLysProProThrAsnProGluSerAsnThrGlnGl 237

Qy   579 CAGCGGCTATGGAGGAGAGGGCTGAGGGAGGAGGCCGAGGAGATGGAGTGCAGGTGGCTGT 638
            ||| ||||:::         ||||||:::
Db   237 nGluAsnLysGlnProGluValGlnLysProGlnLysThrProGluLysPro----- 251

Qy   639 GGTCAGGATCAAGCGCCCCTTGCCCTCCCAG--GTAAACAGATTTACAGAGAAAACTCAA 695
            ||:|||         ::: ||| |:::|||
Db   252 -------LysAlaProLysAlaProGlnGluProGlnAsnProGlyAsnThrGlnAsnAs 268

Qy   696 CTGCAAAGCCCAACACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGGCAGCAGTG 755
               :::|||||
Db   268 nAlaAsnAspGlnGlyValGlnThrGlnGluAspGluIleLysGlnValGlnLysTrpTrpTrpGlnAsnSe 288

Qy   756 GGCTGATGAACACATACA--------------------TCCCAAGCTTCCAAT---CC 791
            |||:::||||         :::                |||:| |||
Db   288 rGlyAlaHisGluValGluLeuThrLysLysIleSerSerAspGlnAsnLeuAsnLeuSe 308

Qy   792 TTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGGCCATCAGCGCGCCCTACACAAGGAGGATGA 851
```

```
Db   308 rPheSerGluGluLeuProGlnGlyIleThrAlaLysLeuValPheSerLysThrAspAs 328
             |||::|
             |||||||||||
Qy   852 GGGCTATGGCCGCCCCAAAGAAGGAGGAACCAAAACTGTGAAGGGCAAGCGTGCTGAGGA 911
             |::: ||||||
Db   328 pLysSerGlyGlnThrLysGluAspPheSerThrThrLysSerLys------------ 344

Qy   912 GCACATCTACAGGGAAATGATGACATGTGCTTCATTATCTGCACAATGGCTCCGCCACAG 971
             ------------------
Db   344 ------------------------------------------------------------ 344

Qy   972 ACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATACGTTCGTATTTC 1031
            :: |||| ||:   ||||||||
Db   345 ---AspGlnLysIleAsnLeu-----GlyGluLeuArgSerGlyLysTyrThrIleSe 361

Qy  1032 AGATAAAGTAGTGGGCATTCTCATG 1056
             ::: ||||||
Db   361 rHisIleIlePheGlyIleAlaLeu 369
```

RESULT 5
JC7862
eukaryotic initiation factor, eIF3 subunit, p116 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 18-Nov-2002 #sequence_revision 18-Nov-2002 #text_change 18-Aug-2003
C;Accession: JC7862
R;Koyanagi-Katsuta, R.; Akimitsu, N.; Hamamoto, H.; Arimitsu, N.; Hatano, T.; Sekimizi
J. Biochem. 131, 833-837, 2002
A;Title: Embryonic lethality of mutant mice deficient in the p116 gene.
A;Reference number: JC7862; MUID:22035555; PMID:12038979
A;Accession: JC7862
A;Molecule type: mRNA
A;Residues: 1-803 <KOY>
C;Comment: This protein, a subunit of the eukaryotic translation initiation eIF3, hav
of mouse development.
C;Genetics:
A;Gene: p116
C;Superfamily: translation initiation factor 3, subunit b (eIF-3b)

Alignment Scores:
Pred. No.:   0.178   Length: 803
Score: 124.00   Matches: 89
Percent Similarity: 32.54%   Conservative: 48
Best Local Similarity: 21.14%   Mismatches: 147
Query Match: 6.03%   Indels: 137
                                 Gaps: 20

US-10-644-659A-1 (1-1146) x JC7862 (1-803)

```
Qy    24 CGGGGAGGGGCCCAGCCAAGAGCGCCCTCCGGAAGAGTACGGCACAGCCACCCTGGTCATCAG 83
             ||  |:::||||:                  ::: |||||
Db    21 ArgGlnGlnProAlaSerGluSerProProThrAspGluAlaAlaGlySerGly------- 38

Qy    84 CTTGGCCCGAGGTTGCAGCAGTGGGCAGCAGAACAGCAT---CAGGCAGGCCCAGGA 140
             :::||||||||||          ::: |||| ||||||       ||||||||
Db    39 ---GlySerGluValGlyLnThrGluAlaProAlaGluGluAspAlaGlyPro-Gl 57

Qy   141 GCCT---ACAGGCTGCCGCGGAGGAGGACCCCAGGAGTCACCTCAAGCTCCTAAACCAAT 197
             ||||      :::         ::|||
Db    57 uProGluValArgAlaLysProAlaAlaGlnSerGluGluThrAlaThrSerProAl 77

Qy   198 CACACCCCTACTTCACACAGAAAGTCAGAGTGCCCCAAAGTCGCCACCCCGCCTGCC 257
             |||||||||||||          ::: |||| |||||
Db    77 aAlaSerProThrPro---GlnSerAlaGluArgSerProSerGluProSerAlaPr 96

Qy   258 A--------------------------------------------------------- 258
Db    96 oGlyLysAlaGluAlaValGlyValGluGlnAlaArgGlyHisProSerAlaGlyAlaGluGl 116

Qy   259 -GAAGGACATGAGATGGACAAAGCTCAGACAAAGCCCTGAGGTTTCTCACATCAAAAA 317
             ||||         :::::         |||||
Db   116 uGluGlyGlySerAspGlySerAlaAlaGluAlaGluProArgAla------LeuGluAs 134

Qy   318 GAAAGAGGTGTCCAAAACGGTGGTCCAGACTTACGACGACTTACGAGAGAGGAGGGGACTGAGCCA 377
```

A;Accession: A24107
A;Molecule type: DNA
A;Residues: 1-838 <SUG>
A;Cross-references: GB:X03346; NID:g21784; PIDN:CAA27052.1; PID:g736319
A;Experimental source: cv. Yamhill
C;Comment: Glutenins, like gliadins, are high in glutamine and proline but differ in
C;Superfamily: glutenin
C;Keywords: seed; storage protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-838/Product: glutenin, HMW chain #status predicted <MAT>

Alignment Scores:
Pred. No.:   838
Score: 0.213   Length: 838
Percent Similarity: 123.00   Matches: 75
Best Local Similarity: 34.90% Conservative: 29
Query Match: 25.17%   Mismatches: 98
DB: 5.98%   Indels: 96
                  1   Gaps: 13

US-10-644-659A-1 (1-1146) x EEWTHW (1-838)

```
Db  134 nGlyGluAlaAspGluProSerPheSerAspProGluAspPheValAspAspValSer--  153
             :::             |||
Qy  378 CCTCAGCCACAGGTACAGAGAGGATGCTGGTCGTGTGCTTGAACCTGGCAGCCAGAAGTGA  437
                               |||       ||||||:::
Db  154 -------GluGluGluLeuLeuGlyAspValLeuGlyAspArgProGlnGluAlaAs  170

Qy  438 C---ATTGACAGATCCTCCACAGCCACGGC------TCCCCAACCCGGAGGAG  482
             :||||  |||||                |||  |||
Db  170 pGlyIleAspSerValIleValAlaValAlaAspAsnValProGlnValGlyProAspArgLeuGl  190

Qy  483 AAAATGTGCCAACCTGGTGTCTGAGCTA----------  510
        ||| ||:::::::
Db  190 uLysLeuLysAsnValIleHisLysIlePheSerLysPheGlyLysIleIleAsnAspTy  210

Qy  511 --ACCAAGGGCTGGAGTGATGGAGCAGGAGGAGCCCAC  548
        :: ||||||:::   |||||||:::
Db  210 rTyrProGluGluAspGlyLysThrLysGlyTyrIlePheLeuGluTyrAlaSerProAl  230

Qy  549 ATGGAGGAGTGACACGTAGACACAGAGACAGGCGGCTATGGAGGAGGCTGAGGAGGAG  608
                   ||||||  |||||   |||
Db  230 aHisAlaValAspAlaValLysAsnAlaAsp---GlyTyr--------------  242

Qy  609 GCCCGAGCAGGATGGAGTGCAGGTGGCTGTGGTTCAGGATCAAGCGCCCCTTGCCCTCCCA  668
                         :::   ||||| :::
Db  243 --------------------LysLeuAspLysGlnHisThrPheAr  251

Qy  669 GGTAAACAGATTTACA----------------  701
        :||||||    |||||    |||
Db  251 gValAsnLeuPheThrAspPheAspLysTyrMetThrIleSerAspGluTrpAsplePr  271

Qy  702 AGCCCAACAGAGAAATATAGCCCAGTGGCAACTTGAAAGGGAGAGTGGCAGCAGTGGGCTGA  761
         :::||| :::              ||||||||
Db  271 oGluLysGlnProPheLysAspLeuGlyAsnLeuArg---------TyrTrpLeuGl  287

Qy  762 TGAACACATACAATCCCGAGAAGTCAATCCTTTCAGTGAAGAGTTTGATTACCAGAGCTGGC  821
                                |||  :::||::::       ||:::
Db  287 uGlu----------AlaGluCysArgAspGlnTyrSerValIlePheGlu--------  300

Qy  822 CATGTCCACCCGCCTACACAAAGAGATGAGGGC-------TATGGCGCCCCAAAGA  872
                :::       |||  |||||||:::
Db  301 -----------------SerGlyAspArgThrSerIlePhePheAsnAspValLysAs  314

Qy  873 AGGAACCAAAACTGCTGAAAGGGCCAAGCGTGTCTGAGGAGCACATCTACAGGGAAATGAT  932
             :       |||||||||:::
Db  314 pProValSerIleGluGluGlnArgGlnArgTrpThrGlu-----------------  326

Qy  933 GGACATGTGCTTCATTATCTGCACAATGGCTCGCCACAGACGATGGCAAGATCCAGGT  992
            :::  :::   ||||||||||
Db  327 ----------------------ThrTyrValArgTrpSerProLysGlyThrTyrLeuAl  339

Qy  993 TACTTTT------------------GGAGATCTCTTGACAGATACGTTCG  1025
        |||||||                :::
Db  339 aThrPheHisGlnArgGlyIleIleAlaLeuTrpGlyIleAlaLeuGlnInleGlnInAr  359

Qy  1026 TATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACCTGTAGACTT  1085
                                              |||
Db  359 gPheSerHisGlnGlyVal-----------------GlnLeuIleAspPh  370

Qy  1086 T 1086
        |||
Db  370 e 370
```

RESULT 6
EEWTHW
glutenin, high molecular weight chain precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 16-Jul-1999
C;Accession: A24107
R;Sugiyama, T.; Rafalski, A.; Peterson, D.; Soll, D.
Nucleic Acids Res. 13, 8729-8737, 1985
A;Title: A wheat HMW glutenin subunit gene reveals a highly repeated structure.
A;Reference number: A24107; MUID:86093674; PMID:3001648

```
Qy    9 GGGCGAAAAGGAAAGCGGGGAGGGGCCCAGCCAAGAGGCGCCCTCCGGAAGATACGGCCACAGC  68
        |||:::::||| |||||||||:::    :::::
Db  464 GlyGlnLysGlyGlnGlnProGlyGlnGlnProGlyGlnGlnGlyGln  483

Qy   69 CACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAACAGCATCAG  128
        |||||||                                     |||
Db  484 GlnProGly                                    GlnGlyGln  489

Qy  129 GCAGGCCCAGGAGCCT---------------------ACAGGCTGGCTGCGGGAGGAGACCCAGGACTC  176
        |||||||                              |||||||
Db  490 -GlnGlnGlnGlnProGlyGlnGlnGlnTrpGlnGlnProGlnGlnSe  509

Qy  177 ACCTCAGCTCCTAAACCA--------------ATCAC  200
        ||| |||                        |||
Db  509 rGlyGlnGlnGlnGlyGlnInTrpGlnGlnProGlyGlyTyrTy  529

Qy  201 ACCCCTACTTCACACCAGAAAGTCAGAGTGCCCAAAGTCGCCACCCGCCTGCCAGA  260
         :::    ||||||                              ||||||
Db  529 rProThrSerProLeuGlnProGlyGlnProGlyGlnTyrAspProThrSerProGl  549

Qy  261 AGGACAGTGAGATGACAAAGCTCAGAGAAGAAAGCCCCTGAGGTTCCACATCAAAAAGAA  320
            :::                                         :::
Db  549 nGlnProGlyGlnGlyGln--------------Pro--GlyGlnLeuGlnGlnProAla  564

Qy  321 AGAGTGTCCAA-------------------AACGGTGGTCAGCAAGACTTACGAGAG  359
        :::|||||                     ||||||    :::
Db  565 GlnGlyGlnGlnGlyGlnGlnLeuAlaGlnGlyGlnGlnGlyGlnGlnProAlaGlnVal  584

Qy  360 AGGAGGGGACGTGAGCCACCTCAGCCACAGGTA------------CGGAGAG  398
        |||||||                                     |||:::
Db  585 GlnGlyGlnGlyGlnProAlaGlnGlyGlnGlnGlnGlyGlnLeuGlyGlnGln  604

Qy  399 GGATGCTGGTGTGTGCTTGAACCTGGGCAGCAGCCAGAATGACATTGACGAGAATCCTCCACAG  458
            |||                                         ||||||
Db  605 Gly--------------------------------------GlnGlnProGlyGlyGln  610

Qy  459 CCACGGCTCCCAACGCGGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGG  518
            :::||:::::                              |||||
Db  611 GlyGlnInProAlaGlnGlyGlnGlyGlnGlnProGlyGlnGly  624

Qy  519 CTGGAGAGTGATGGAGCAGGAGGAGGAGCCCACATGGAGGAGGCTGAGGAGGTGACAGCGCACACAGAGGA  578
                                              |||:::   :::||:::
Db  625 ---------------GlnGlyGlnGlnGlyGlnQlnProGlyGlnnProGlyGlnGly  641

Qy  579 CAGGGGCTATGG---------------------------AGGAGAGGCTGAGGAGAGCGCCCCAGCAGGA  620
        :::                                   |||:::
Db  642 GlnProTrpTyrProTyrProThrSerProGlnGluSerGlyGlnGlnGlyGlnGlnProGlyGlnGln  661

Qy  621 TGGAGTGCAGGTGGC---------------------TGTGGTCAGGATCAA  650
        |||                                :::
Db  662 TrpGlnGlnProGlyGlnTrpGlnGlnProGlyGlnProGlyTyrTyrLeuThr  681
```

Qy 651 GCGCCCCTTGCCCTC---CCAGGTAAACAGATTTACAGAGAGAAACTGCAAAGCCCA 707
        :::|||||| |||:||| ||||||| |||| |||| || ||||||
Db 682 SerProLeuGlnLeuGlyGlnGlyGlnGlyGlyGlyThrTyrProThrSerLeuGlnGlnPro 701

Qy 708 ACAGAAAATATAGCCCAGTGGCAACTTGAAAGGGAGATGGCAGCAG 753
        ::: ||||||||||
Db 702 GlyGln-----------GlyGlnProGlyGlyGlnTrpGlnGln 712

RESULT 7
S70642
ubiquitin ligase Nedd4 - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 14-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S70642
R;Staub, O.; Dho, S.; Henry, P.C.; Correa, J.; Ishikawa, T.; McGlade, J.; Rotin, D.
EMBO J. 15, 2371-2380, 1996
A;Title: WW domains of Nedd4 bind to the proline-rich PY motifs in the epithelial Na(+)
A;Reference number: S70642; MUID:9221297; PMID:865844
A;Accession: S70642
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-887 <STA>
A;Cross-references: EMBL;U50842; NID:g1293646; PIDN:AAB48949.1; PID:g1293647
C;Genetics:
A;Gene: Nedd4
C;Superfamily: rat ubiquitin-protein ligase; protein kinase C C2 region homology; ubiqui
F;54-167/Domain: protein kinase C C2 region homology <KC2>
F;246-283/Domain: WW repeat homology <WW1>
F;402-439/Domain: WW repeat homology <WW2>
F;459-496/Domain: WW repeat homology <WW3>
F;555-881/Domain: ubiquitin-protein ligase homology <UBI>

Alignment Scores:

Pred. No.:   0.255   Length: 887
Score: 122.00   Matches: 86
Percent Similarity: 35.85%   Conservative: 66
Best Local Similarity: 20.28%   Mismatches: 128
Query Match: 5.93%   Indels: 144
DB: 2   Gaps: 21

US-10-644-659A-1 (1-1146) x S70642 (1-887)

Qy 34 CCAGGAAGAGCGCCCTCCGGAAGATACGCACAGCCACCCTGGTCATCAGCTTGGCCCGA 93
       |||
Db 107 ProMetSerGlyValLeuThrSerValGlnThrLysThrLysSerLeuAsnPro 126

Qy 94 GGTTGGCAGCAG--105
       ||| :::
Db 127 LysTrpAsnGluGlulleLeuPheArgValLeuProGlnHisHisArgIleLeuPheGlu 146

Qy 106 ----------------------TGGGCGAATGAGAACAGCATCAGCGCCAGGCCCAGGAGCCTACA 147
        ::::|| ||||
Db 147 ValPheAspGluAsnArgLeuThrArgAspAspPheLeuGlyGlnValAspValProLeu 166

Qy 148 GGCTGGCTGCCGCCGGAGGGACCCAGGACTCACCTCAAGCTCCTAAACCAATCACA-----201
        ||||| ||||:||:::
Db 167 TyrProLeuPro--------ThrGluAsnProArgMetGluArgProTyrThrPheLys 183

Qy 202 --------CCCCTACTTCACACCAGAAAGTCAGAGTGCCCCAAAGTCGCCACCC 249
        ||||| |||
Db 184 AspPheValLeuHisProArgSerHis--------LysSerArgValLysGlyTyrLeu 200

Qy 250 CGC------CTGCCAGAAGGACATGGAATGGACAAAGCTCAGAGAAAGCC 294
        ||| ::: |||||||| :::||||||
Db 201 ArgLeuLysMetThrTyrLeuProLysAsnGlySerAspAspGluAsnAlaAspGlnAla 220

Qy 295 CCTGAGGTT------------------TCTCACATCAAA 315
        ||:::   ::|||||:|:::
Db 221 GluGluLeuGluProGlyTrpValValLeuAspGlnProAspAlaAlaThrHisLeuGln 240

Qy 316 AAGAAA------------------------GAGGTGTCCAAAACGGTGGTCAGC 345
        ||||| ::::|||||:::
Db 241 HisProGluProSerProLeuProProGlyTrpValProProGlyTrpGluArgGlnValAspGln 260

Qy 346 AAGACTTACGAGAGAGGAGGGGACGTGAGCGCACCTCAGCCAC-----AGGTACGAGAGG 399
        :::|||||| || |||||| |||||
Db 261 ArgThrTyr-------------ValAsnHisGluSerArgThrArgThrGlnTrpLysArg 276

Qy 400 GATGCTGGTGTGCTTGAACCTGGCCAGCCAGCAGAATGACATT------------------441
        ||| :::
Db 277 ProSer-----------------------------ProGluAspLeuThrAspAspGluAsnGly 289

Qy 442 GACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTG 501
        ||| :::|||||||::: ||| ||||| :::
Db 290 AspIleGlnLeuGlnAlaHisGlyAlaPheThrThrArgArgGlnIleSerGluAspVal 309

Qy 502 TCT---------------GAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAGGAGCCC 546
        |||
Db 310 AspGlyProAspAsnHisGluSerProGluAsnTrpGluHisValArgGluAspGluAsn 329

Qy 547 ACATGGAGGAGTGACAGCGTAGACACAGAGGACACGGCTATGAGGAGGCTGAGGAG 606
        ||| |||:|||
Db 330 ThrIle-------------------------------------TyrSerGlyGlnAlaValGln 338

Qy 607 AGGGCCGGAGCGAGGATGGAGTGCCAGGTGGCTGTGTCAGGATCAAGCGCCCCTTGCCCTCC 666
        |||||| |||||||||||| ||||||:::
Db 339 SerProProSerGlyHisProAspValGln---ValArgLeuAlaGluGluLeuLeuAspThr 357

Qy 667 CAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCC-----AACAGAAATATAGC 720
        ::::|||||||:::|||:::||||||||
Db 358 ArgLeuThrMetTyrGlyAsnPro-AlaThrSerGlnProValThrSerSerAsnHisSe 377

Qy 721 CCAGTGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACACATAACAATCCCAG 780
        ::|||| |||:::
Db 377 rSer-----------ArgThrGlySerSerGlnThrCysIlePheHisGluGluGlnProTh 393

Qy 781 AAGCTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCACCC--------832
        ::|||
Db 393 rLeuProValLeu--------LeuProThrSerSerGlyLeuProProGlyTrpGl 409

Qy 833 GCCTACACAAAGGAGGATGAGGGGCTATGGCCGCCCC 867
        ::||| :::
Db 409 uGluLysGlnAspAspArgGlyArgGlyArgGlySerTyrTyrValAspHisAsnSerLysThrThrTh 429

Qy 868 AAAGAAGGAACCAAAACTGCTGAAAGGGCCA----AGCGTGCTGAGGAGCACATTCACAGG 924
        ::||| :::
Db 429 rTrpSerLysProThrMetGlnAspAspProArgSerLysIle-ProAlaHisLeuArgG 449

Qy 925 GA--926
        ||
Db 449 lyLysThrProValAlaSerAsnAspLeuGlyProLeuProProGlyTrpGluArgT 469

Qy 927 ----AATGATGGACATGTGCTTCATTATCTGCACAATGGCTCGCCCAGAGCGAGATGGC 981
        ||| :::
Db 469 hrHisThrAspGlyValPhePheIleAsnHisAsn---IleLysLysThrGlnTrpG 488

Qy 982 AAGATCCA 989
        |||
Db 488 luAspPro 490

RESULT 8
I48378
hairless protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
C;Accession: I48378
R;Cachon-Gonzalez, M.B.; Fenner, S.; Coffin, J.M.; Moran, C.; Best, S.; Stoye, J.P.
Proc. Natl. Acad. Sci. U.S.A. 91, 7717-7721, 1994
A;Title: Structure and expression of the hairless gene of mice.
A;Reference number: I48378; MUID:94329587; PMID:8052649
A;Accession: I48378
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1182 <RES>
A;Cross-references: EMBL;Z32675; NID:g531706; PIDN:CAA83587.1; PID:g531707

Alignment Scores:
Pred. No.:
Score: 0.313
Percent Similarity: 35.51%
Best Local Similarity: 24.80%
Query Match: 5.88%
DB:

Length: 1182
Matches: 95
Conservative: 41
Mismatches: 117
Indels: 130
Gaps: 24

US-10-644-659A-1 (1-1146) x I48378 (1-1182)

Qy   23 GCGGGGAGGGGCCAG----------------------------------CCAAGA 43
         ||||||||| |||| |||||||
Db 255 AlaGlyArgGlnAsnLeuCysProValPheGlyLeuTyrProAspThrValProArg 274

Qy   44 GCGCCCTCCGGAAGATACGCACAGCCACCCTCGT------CATCAGCTTGGCCCGAGGTT 97
         ::: |||||| ||||| ||||||
Db 275 AlaProTrpPro-------SerCysPro-ProGlyLeuValHisSerLeuGly------ 289

Qy   98 GGCAGCAGTGGGCGAATGAGAACAGCATCAGGCAGGCCAGGAGCC---------TA 145
                                                ||||||||||||||||||||||
Db 290 ---------------------------AsnIleTrpAlaGlyProGlySerAsnSerLeuGlyTy 302

Qy 146 CAGGCTGCTGCCGGGAGGACCAGCTCACCTCAAGCTCCTAAACCAATCACACCCC 205
         |::|||| | :::||| ||||||||
Db 302 rGlnLeuGly-ProProAlaThrProArgCysPro---SerProGlyProProThrProP 321

Qy 206 CT----ACTTCACCACCAGAAAGCTTCAGAGTGCCCCAAAGTCGCCACCCCGC- 252
         | :::||||||
Db 321 roGlyGlyCysCysSerSerHisLeuProAlaArgGluGlyAspLeuGlyProCysArgL 341

Qy 253 --------CTGCCAGAAGGA------CATGGAGATGGACAAAGCTCAGAGAAA--- 291
                  |||||||||| :||||||||||||||||||||
Db 341 ysCysGlnAspSerProGluGlyGlySerGlyProGlySerSerGlyGluArgA 361

Qy 292 ---------GCCCCTGAGGTTTCTCACATCAAAAAGAAGAGGTGTCCA 331
                   |||
Db 361 snLysAlaAspSerArgCysProProSerHisHisThrLysLeuLysLysThrTrpL 381

Qy 332 AAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACGTGAGCCACCTCAGCCACAGGT 391
         :::|||| |||||| |||||||||||||||
Db 381 euThrArgHisSerGluGlnPheGluCysProGlyGlyCysSer----------- 395

Qy 392 ACCAGAGAGGAGGATGCTGGTGTGCTTGAACCTGGGCACCCAGGAAGAATGACATTGAC------A 445
                                                  ||:::: |||:||
Db 396 ----------------------------GlyLysGluSerSerAlaThrGlyLyleuA 406

Qy 446 GAATCCTCCACAGCCACGGCTCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTTCTG 505
         |||||||||||||| ||||| :::
Db 406 rgAlaLeuLysArgAlaGlySerProGlu-----------------------ValGlnG 418

Qy 506 AGCTAACCAAGGCTGGAGAGTGGATGGAGTGAGAGGAGGGCCC---------- 546
         |||
Db 418 lyAlaSerArgGlyProAlaProLysArgProSerHisProGlyThrGlyArgG 438

Qy 547 ----------ACATGGAGAGTGCACAGCGTAGACACGAGGACAGC-----GGCTATG 589
         ||:::   :::||||
Db 438 lnGlyAlaArgAlaTrpGln-------------------GluThrProGluThrIleIleIleGlySerL 454

Qy 590 GAGGAGAGGCTGAGGAGGAGGCCCGAGCAGCAGGATGAGTGAGGTGGCTGTGGGTCAGGATCA 649
         |||||||||||||| |||||||| ||| ||
Db 454 ysAlaGluAlaGluGluGluGluGlnLeuArgAspArgAspArgAspGlyArgIleArgLeu-- 473

Qy 650 AGCGCCCCCTGCCCTCCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAA-------- 702
         :::|||||| ::: ||||||||||
Db 474 ----------------GlnSerArgLeuValAsp---ThrSerCysGlnHisHisL 487

Qy 703 ------------------------GCCCAACAGAAATATAGCCCAGTGGGCAACTTGA 736
         :::||||||
Db 487 euAlaGlyValThrGlnCysGlnSerCysValGlnAlaAlaGlyGluValValGlyValLeuT 507

Qy 737 AAGGAGATGGCGCAGCAGTGGGCTGATGAACACATACAATCCCAGAAGCTCAATCCTTTCA 796
                                   ||| :::|||||
Db 507 hrGly--------------------HisSerGlnLysSerArgArgSerProLeuG 519

Qy 797 GTGAA---GAGTTTGATTACGAGTGCCATGTGGCCATGTCCACCGCCTACACAAAGGAGATGAGG 853
         || ::: ||| ::: :::|||
Db 519 luGluLysGlnLeuGluGluGluGlnAspSerSerAlaThr-------SerGluGluG 535

Qy 854 GCTATGGCGCGCCCCAAAGAAGAGAACAACAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGC 913
         ||| |||:::|||
Db 535 lyGlyGlyPro-----------------------------GlyProGluAlaSerLeuAsnLysGlyLeuAlaLysH 552

Qy 914 ACATTACAGGAAAATGATGATGGACAATGTGCTTCATTATCTGCACAACTGGTCTCGCCACAGAC 973
         ||| ::: |||:||| |||
Db 552 isLeuLeuSerGlyLeuGlyAspArg-----------------LeuCysArgLeuLeuArgGlysGluA 569

Qy 974 GAGAT 978
         ||:|||
Db 569 rgGlu 570

RESULT 9
S02262
glutenin high molecular weight chain Dx5 - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 03-Feb-1994
C;Accession: S02262
R;Anderson, O.D.; Greene, F.C.; Yip, R.E.; Halford, N.G.; Shewry, P.R.; Malpica-Romei
Nucleic Acids Res. 17, 461-462, 1989
A;Title: Nucleotide sequences of the two high-molecular-weight glutenin genes from th
A;Reference number: S02262; MUID:89098819; PMID:2563152
A;Accession: S02262
A;Molecule type: DNA
A;Residues: 1-848 <AND>
A;Cross-references: EMBL:X12928
C;Genetics:
A;Gene: Glu-D1-1b
C;Superfamily: glutenin

Alignment Scores:
Pred. No.:
Score: 0.329
Percent Similarity: 120.50
Best Local Similarity: 35.34%
Query Match: 24.11%
DB: 5.86%

Length: 848
Matches: 88
Conservative: 41
Mismatches: 131
Indels: 105
Gaps: 15

US-10-644-659A-1 (1-1146) x S02262 (1-848)

Qy   9 GGGCGAAAAGGAAAGCGGGGAGGGCCCAGCCAGAAGACGCGCCCTCCGGAAGATACGCACAGC 68
         ||:::||||||||||:::|||| |||:::
Db 459 GlyGlnLysGlyGlnGlnProGlyGlnGlnProGlyGlnGlyGlnGln 478

Qy   69 CACCCTGGTCATCAGCTTGGCCCGAGGTTGGCCAGCAGTGGGCGAATGAGAACAGCATCAG 128
         |||| ---GlnGlyGln 484
Db 479 GlnProGly--------------------------------GlnGlyGln 484

Qy 129 GCAGGCCCAGGAGCCT----------ACAGGCTGCTGCCGGGAGGGACCCAGACTC 176
         ||:::||| |||
Db 485 -GlnGlyGlnGlnProGlyGlnGlnProGlyGlnTyrTyrProThrSerProGlnGlnSe 504

Qy 177 ACCTCAAGCTCCTAAACCA------------ATCAC 200
                       ::||||
Db 504 rGlyGlnGlyGlnGlnProGlyGlnGlnTrpGlnGlnProGlyGlnGlyProGlyTyrTyrTy 524

Qy 201 ACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCAAAGTCGCCACCCGGCCAGA 260
         :::|||||||||||||||
Db 524 rProThrSerProLeuGlnProGlyGlnGlnTrpGlnProGlyGlnProGlyTyrTyrAspProThrSerProGl 544

Qy 261 AGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAA 320
                                            ---Pro--GlyGlnLeuGlnGlnProGlyGlnProAla 559
Db 544 nGlnProGlyGlnGlyGln---------------Pro--GlyGlnLeuGlnGlnProGlyGlnProAla 559

Qy 321 AGAGGTGTCCAAAACCGGTGGTCAGCAAGACTTACGAGAGAGGAGGGGGACGTGAGCCACCT 380
         ||| GlyGlnGlnLeuAlaGlnGlnGlyGlnGlnGlyGlnGlnProAla 577
Db 560 GlnGlyGlnGln----------GlyGlnGlnLeuAlaGlnGlnGlyGlnGlnGlyGlnGlnProAla 577

Thu May 6 09:39:11 2004   us-10-644-659a-1.n2p.rpr

```
Percent Similarity:      32.65%         Conservative:    31
Best Local Similarity:   24.68%         Mismatches:     139
Query Match:              5.86%         Indels:         124
DB:                           2         Gaps:            21

US-10-644-659A-1 (1-1146) x T42761 (1-3938)
```

```
Qy      23 GCGGGGAGGGCCCAGCCAAGAGCGCCCTCCGGA-----AGATACGCACAGCCA-            70
              ::: |||||||         :::  |||||
Db    3031 AlaSerAlaAlaValProAlaThrProSerGlyProThrAlaPheGlnProArgPhe      3050

Qy      71 CCCTGGTCATCAGCTTGGCCCGAGGTTGCAGCAGTGGCG-AATGAGAACAGCATCAGG     129
              ::::::| ||||||||||||||:: |||:|
Db    3051 ProProAlaAlaThrGlnTyrArgAlaGlySerSerGlyProThrGlnAsnGlyPheLeu   3070

Qy     130 CAGGGCCCAGAGCCTACAGCTGCTGGCTGCGTCCGGGAGGACCCAGGACTCACCTCAAGCTCCT    189
              ||||| ||||| |||||||||||||::  :::  :::
Db    3071 AlaHisGlnAlaProThrTyrThrGlyPro---SerThrTyrProAlaProThrTyrPro     3089

Qy     190 AAACCAATCACACCCCCTACTTCACACCAGAAAGTCAGAGTGCCCAAAGTCGCCACCC     249
              |||||||||:::                   ||||||
Db    3090 ---------------ProGlyThrSerTyrProAlaGluProGlyLeuProSerGlnProAla   3105

Qy     250 CGCCTGCCAGAAGGACAT                                               267
              |||||||||||||
Db    3106 PheHisProThrGlyHisTyrAlaAlaProThrProMetProThrThrGlnSerAlaPro   3125

Qy     268 -----------------GGAGATGGACAAAGTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAA    315
              :::  :::  |||:::
Db    3126 PheProValGlnAlaAlaAspSerHisAlaAlaHisGlnLysProArgGlnThrSerLeuAla   3145

Qy     316 AAGAAAGAG-GTG                                                    327
              |||
Db    3146 AspLeuGluGlnLysValProThrAsnTyrGluValIleSerSerProAlaValThrVal   3165

Qy     328 TCCAAAACG-----------------GTGTCAGCAAGACTTACGAG-                 357
              ||||||||                   ||||||||||||:::
Db    3166 SerSerThrProSerGluThrGlyThrTyrSerGlyProAlaValSerSerTyrGluHis   3185

Qy     358 -------------AGAGGAGGGGACGTGAGCCACCTCAGCCAC-                    387
              |||||||||||||
Db    3186 GlyLysAlaProGluHisProArgGlyValAspArgSerVerSerGlnSerProAla     3205

Qy     388 ---AGTACGAGAGGATGTCGGTGTG-------------CTTGAACCTGGGCAGCCAGAGAATGAC    438
              |||||||||||||    :::::  |||||
Db    3206 ProThrTyrProSerAspSerHisTyrThrSerLeuGluGlnAsnValProArgAsnTyr   3225

Qy     439 ATTGACAGAATCCTCCACAGCCTCCCACGCGGAGGAGAAAATGCCACCTG            498
              :::
Db    3226 Val----------------------------------------MetIleAspAsp        3230

Qy     499 GTGTCTGAGTCAACCAAGGGGCTGGAGAGTGATGAGAGTGAGGAGCCCACATGGAGGAGT       558
              ::|||||||||||||
Db    3231 IleSerGluLeuThrLys------------------------AspSerThrProThr---AlaSer    3243

Qy     559 GACAGGCGTAGACACAGAGACAGCCGACCGGCCGACCGGCCCGAGCAGGCCCGAGGAGGCCCGAGCAG    618
              ||||||||
Db    3244 AspSerGlnArgProGluHisProLeuGlyProGlyVal--ValSerGlyArgProGlyLys   3262

Qy     619 GATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCC---TTGCCCTCCCAGTAAAC      675
              |||  ||||||    :::::
Db    3263 AspProGlyGluProAlaVal-----LeuGluGlyLeuProThrCysTyrGlyTyrGly   3280

Qy     676 AGATTTACAGAGAAACTCAACTGACAAGCCCAACAGAAATATAGCCCAGTGGCAACTTG       735
              :::
Db    3281 ArgGlyGluGlu------SerGluGluAspSerTyrAspProArgGlyLys---       3296

Qy     736 AAAGGGAGATGGCAGCAGTGGCTGATGAACACATACAATCCAGAGCTCAAT------       789
              |||  :::  :::
Db    3297 ----------------SerGlyHisHisArgSerMetGluSerAsnGlyArg          3308

Qy     790 CCTTTCAGTGAAGAGTTT-----GATTACGGAGCTGGCCATGTCCACCCGCTACACAAA      843
              |||  |||                 :::  :::            |||
```

```
Qy     381 CAGCCACAGGTACGAGAGGATGTCGGTGTGTCTTGAACCTGGGCAGCCAGAGAATGACAT    440
              |||
Db     578 GlnValGln-----------------------------------------------       580

Qy     441 TGACAGAATCCTCCACAGCCACCGCGCTCCCCACGCGGAGGAGAAAATGTGCCAACCTGGT     500
              ::: |||
Db     581 ---------------GlnGlyGlnArgProAlaGlnIleGlnIleGlnIleGlnProGly     595

Qy     501 GTCTGAGCTAACCAAGGGCTGGAGAGTGATGAGAGAGTGAGGAGCCCACATG-----     551
              |||||||         :::
Db     596 ----------GlnGlyGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGln     611

Qy     552 ----------GAGGAGTGACAGCAGGAGCACAGGAGACAGCGGCTATGGAGGAGGAGGC     599
              :::::  |||||:::|
Db     612 ProGlyGlnGlyGlnGlnGlyGlnGlnProAlaGlnGlyGlnGlnGlnGlyGlnGlnProGly   631

Qy     600 TGAGGAGAGGGCCCGAGCAGGATGGAGTGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTT    659
              ||||:::
Db     632 GlnGlyGlnGlnGlyGlnGlnProGlyGlnGlnGlnGlnGlnGlnGlnProGlyGlnGlyGly    651

Qy     660 GCCCTCCCA-----GGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCCAACAGAA     713
              |||||
Db     652 GlnGlnProGlyGlnGlyGlnGlnProGlnTrpProTyrProThr-------SerPro-GlnGl   668

Qy     714 ATATAGCCCAGTGGGCAACTTGAAAGGGGAGGAGTGCAGCAGTGGCTGATGAACACATACA     773
              :            ||||||||||:::|||||
Db     669 u----SerGlyGlnGlyGlnGlnProGlyGlnTrpProGlnGln-------------     680

Qy     774 ATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGGAGCTGGCCATGTCCACCCG    833
              :::       ||| |||
Db     681 -----------------ProGlyGlnProGlyTyrProTyrLeuThrSerProLeuGl     695

Qy     834 CCTACACAAAGGAGGATGAGGGCTAT------GGCCGCCCCAAAGAAGGAAC         878
              ::| |||:::  :::::|
Db     695 nLeuGlyGlnGlyGlnGlyTyrProThrSerLeuGlnGlnGlnProGlyGlnGly--     714

Qy     879 CAAAACTGCTGAAAGGGCCAAGCGTGCTGAGAGACACATTCACAGGGAAAATGATGG-----     934
              :::::        ||| ::: ||||||
Db     715 -----------GlnGln-ProGlyGlnTrpProGlnSerGlyGlnGlnGlnGlnHisTrpTyrT    731

Qy     935 -----ACATGTGCTTCATTATCTGCACAATGGCTCGCCCAGAGCAGAGATGCCAAGATCCA     989
              |||||
Db     731 yrProThrProSerProGlnLeuSerGlnGlyGln-GlyGlnAsnProGlyGlnTrpLeuGlnPro    750

Qy     990 GGT     992
              |||
Db     751 Gly     751
```

```
RESULT 10
T42761

Bassoon protein - rat
N;Alternate names: brain-specific synapse-associated protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C;Accession: T42761
R;Dieck, S.; Sanmarti-Vila, L.; Langnaese, K.; Richter, K.; Kindler, S.; Soyke, A.; Wex,
J. Cell Biol. 142, 499-509, 1998
A;Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
A;Reference number: Z22249; MUID:98345363; PMID:9679147
A;Accession: T42761
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-3938 <DIE>
A;Cross-references: EMBL:Y16563; NID:g3413503; NID:CAA76287.1; PIDN:CAA76287.1; PID:g3413504
A;Experimental source: strain Sprague Dawley; brain
C;Function:
A;Description: may be involved in cytomatrix organization at the site of neurotransmitte
A;Note: component of the presynaptic cytoskeleton
C;Keywords: coiled coil; zinc finger
```

```
Alignment Scores:
Pred. No.:     0.39        Length:     3938
Score:       120.50        Matches:      96
```

Db   3309 ProAlaSerThrHisTyrTyrSerAspTyrArgHisGlyAlaArgAlaAspLys 3328

Db   206 SerProProArgProArgProSerSerProAlaArgArgLysAspAspAspLeu--- 224
                                                                 :::|||
Qy   844 GGAGATGAGGGCTATGGCCGCCCCAAAGAAGAAGAACCAAAACTCTGAAAGGGCCAAGCGT 903
               |||| |||
Qy   502 TCTGAGCTAACCAAGGGCTGGAAGGTGATGAG---------------CAGGAGGAGCCC 546

Db   3329 TyrGlyProGlyProMetGlyProLysHisProSerLysAsnLeuAlaProAlaAla--- 3347
               ||||| :::|||
Db   225 SerProProArgLysArgLysIleGluProLysLysIleLysLysGluGluPro 244

Qy   904 GCTGAGGAGCACATCTACAGGGAAATGATGGAACATGTGCTTCATTATCTGCACAATGGCT 963
               |||||
Qy   547 ACATGGAGGAGTGAGACACGTAGACACAGAGACACAGCGGCTATGGAGGAGGCTGAGGAG 606

Db   3348 --------------------------------IleSerSerLysArgSer 3353
                                             |||||
Db   245 --------------AspSer---AspThrGluThrGlyArgThrLeuGlyGly------ 257

Qy   964 CGCCACAGACGAGATGGCAAGATCCAG 990
               :::|||||||::: |||
Qy   607 AGGCCCGAGCAGCAGGATGGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCCTCC 666

Db   3354 LysHisArgGlnGlyMetGluGln 3362

Db   258 -----LysArgSerGlyLeuGlnSerAla---------------ArgAspLeuLysGlu 270

RESULT 11
S24457
hypothetical protein R08D7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Sep-1997
C;Accession: S24457; S1036
R;Sulston, J.; Du. Z.; Thomas, K.; Wilson, R.; Hillier, L.; Staden, R.; Halloran, N.; Gr
awkins, T.; Ainscough, R.; Waterston, R.
Nature 356, 37-41, 1992
A;Title: The C. elegans genome sequencing project: a beginning.
A;Reference number: S24457; MUID:92168156; PMID:1538779
A;Accession: S24457
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-458 <SUL>
A;Cross-references: EMBL:Z12017; NID:g6812; PID:g6813
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1992
C;Genetics:
A;introns: 35/2; 145/1; 170/1; 428/2

Qy   667 CAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACACAGAAAATAGCCCAGTG 726
               :::   :::|||||
Db   271 Glu-------------SerAspLysLeuArgAlaLysAsnSerLysMetPheGluGluMet 286

Alignment Scores:
Pred. No.:   0.366
Score: 119.50
Percent Similarity: 38.36%
Best Local Similarity: 23.54%
Query Match: 5.81%
DB:

Qy   727 GGC---AACTTGAAAGGGAGATGGCAGCAGTGGCGTGATGAACACATACAATCCCAGAAG 783
               ||| |||||:::
Db   287 AspThrSerValSerGlyArgPheAlaAspThrValTyrArgGlnLysGlnThrLysLys 306

US-10-644-659A-1 (1-1146) x S24457 (1-458)

Qy   784 CTCAATCCTTTCAGTGAAGAGTTTGATTACGAAGCTGGCCATGTCCACC---CGCCTACAC 840
               ||| |||||:::   :::
Db   307 LysGlyLysAspSerGluGluValAspGlnAlaLysLysGluArgGluThrLysLysThrGlu 326

Qy   115 GAGAACAGCATCAGGCAGGCCAGGAGCCTACAGGCTGCGTGCGGAGGGACCCAG--- 171
               ::: ||||||||||:::|||
Db   68 LysGlnSerValLysLysAlaLysValValHisGlyPheLysGlnThrPheAlaGluVal 87

Qy   841 AAAGGAGTGAGGGCTATGGCCGCCCCAAAGAAGAGAACCAAAACTGCTGAAAGGGCCAAG 900
               ::: ||| :::
Db   327 GluLeuLysGluLysTyrLysSerTrpAsnLysGlyValAlaGlnIleGluAspArgArg 346

Qy   172 GACTCACCTCAAGCT------CCTAAACCAATCACACCCCCTACTTCA---------- 213
               ||||||:::||::: |||
Db   88 AspAlaProLysValIleLeuLysProGluProLeuSerProAspAsnSerProProArgGly 107

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAAATGATGGACATGTGCTTCATTATCTGCACAATG 960
               ||| ::: |||
Db   347 AlaGlnLeuGluGlyGluMetAlaAlaArgValAlaAlaGlu----ProMet 360

Qy   214 --------CACCAGAAAGCTTCAGAGTGCCCCAAAG----------- 240
               |||||||:::
Db   108 LysArgGlnHisAspSerAspAsnSerProProArgProArgLysArgAsnAsp 127

Qy   961 GCTCGCCACAGACGAGATGGCAAGATCCAGCGGTTACTTTGGAGATCTCTTTGACAGATAC 1020
               |||||| :::|||
Db   361 AlaArgAlaArgAspAspAspMetAsnAlaMetAsnAlaHisLeuLysGluVal------ 376

Qy   241 --------TCGCCACCCCGCCTGCGAGAGGACATGGCAGCCAAAGCTCAGAGAAA 291
               ::: ||||||||||||
Db   128 SerAspAsnSerProArgProArgProSerArgAsnArgHisAspSerAspLysAspAspSer 147

Qy   1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGAAAACATGGACTCGGTA 1080
               ::: ||| :::
Db   377 LeuHisIleAlaIleAlaAspProMetAlaAsnMetIleGlnIleLysArgArgAspThrAlaIle 396

Qy   292 GCCCCTGAGGTTTCTCCACATCAAAAGAAGAGAGGTGTCCAAAACGGTGGTCAGCAAGACT 351
               |||
Db   148 ProPro-ArgArgArgHisAspAsnSerProProArgGluAsnArgSerArgArgSerArgLys-I 167

Qy   1081 GACTTTGAAGGAGGAGATGCTATGGCAAGGCCGAGATGGACCATGTTGTG 1128
               |||||||:::||:::
Db   397 Asp--ArgGlyGluLeuValTyrProSerTyrHisGlyHisPheVal 411

Qy   352 TACGGAGAGG----------AGGGGACGTGAGCGAGGCCACCTCAGCCACCAGGTACGAGAGG 399
               :::|||||
Db   167 leArgGluGluSerProSerAlaArgAsnArgArgSerPro---ProArgThrArg-Arg 185

RESULT 12
B30843
glutenin high molecular weight chain Ax2 precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 15-Nov-1996
C;Accession: B30843
R;Anderson, O.D.; Greene, F.C.
submitted to GenBank, January 1989
A;Reference number: A94515
A;Accession: B30843
A;Molecule type: DNA
A;Residues: 1-815 <AND>
C;Superfamily: glutenin
C;Keywords: seed; storage protein
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-815/Product: glutenin Ax2 chain #status predicted <MAT>
F;108-773/Region: glutamine/glycine/proline-rich

Qy   400 GATGTCTGGTGTGCTTGAACCTGGGCAGCCAGAGAAT----------GACATT 441
               ||| ||| :::
Db   186 AspArgHisAspSerAspAsnSerProProArgAsnArgSerArgArgArgSerAspAsn 205

Alignment Scores:
Pred. No.:   0.464
Score: 118.50
Percent Similarity: 33.97%
Best Local Similarity: 24.66%
Query Match: 5.76%
DB:

Qy   442 GACAGAATCCTTCCACAGCCACGGCGTCCCCCAACGCGGGAGAGAGAAAATGTGCCAACCTGGTG 501

Length: 815
Matches: 90
Conservative: 34
Mismatches: 131
Indels: 18
Gaps: 2

Length: 458
Matches: 89
Conservative: 56
Mismatches: 161
Indels: 72
Gaps: 19

Qy   980 GCAAGATCCAGGT 992
          ||||| ||||
Db   629 pGlnGlnProGly 633

RESULT 13
S15720
glutenin high molecular weight chain 1Ax1 precursor - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 08-Jun-1994 #sequence_revision 12-May-1995 #text_change 20-Aug-1999
C;Accession: S15720
R;Halford, N.G.; Field, J.M.; Blair, H.; Urwin, P.; Moore, K.; Robert, L.; Thompson,]
submitted to the EMBL Data Library, July 1991
A;Description: Analysis of HMW glutenin subunit encoded by chromosome 1A of bread whe.
A;Reference number: S15720
A;Accession: S15720
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-830 <HAL>
A;Cross-references: EMBL:X61009; NID:g21742; PIDN:CAA43331.1; PID:g21743
C;Superfamily: glutenin

Alignment Scores:
Pred. No.:   0.553   Length: 830
Score: 117.50   Matches: 88
Percent Similarity: 32.28%   Conservative: 35
Best Local Similarity: 23.10%   Mismatches: 126
Query Match: 5.71%   Indels: 132
DB: 2   Gaps: 19

US-10-644-659A-1 (1-1146) x S15720 (1-830)

[sequence alignment rows — Qy/Db pairs]

US-10-644-659A-1 (1-1146) x S74544 (1-1001)

```
Db   516 |||:::             |||:::              |||
         GlnGlnPro-------------GlyGlyGlyGlyTyrTyrProThrSerPro 530

Qy       12 CGAAAAGAAAGCGGGAGGGGAGGGCCCAGCCAAGAGCGC---     CCTTCCGGAAGATACG  62
Qy   621 -----------------------TGGAGTGCAGTGGCTGTGGTCAGGAT 647

Db       89 LysGlnGluLysSerGlyGlyLnTrpGlnGlnSerGlyGlnGlyGln---
            LysGlnGluLysSerGlyGlyLnProSerProAlaArgProThrProProArgProLys 108

Qy   531 GlnGlnSerGlyGlyGlnGluGlnLnLeuGlnGluGlnLnTrpGlnGlnSerGlyGlnGlyGln---  549
Qy       63 CACAGCCACCTGGTCATCAGTGTCCCGAGG--TTGGCAGGAGTGGGCAGATGAGAA  119

Db       109 LeuGlnAlaProLysAlaProThrProProGlnProProValAlaLys-AlaSerAlaPr  128

Qy   648 CAAAGCGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAA------  701
Qy       120 CAGGATCAGGCAGGCCCCAGGAGCCTACAGGCTGCTGCCGGGAGGGACCCAGGACTCACC  179

Db       550 ProGlyHis------TyrProThrSerProLeuGlnProGly  561
Db       128 oLysIleGlnLysGlnGluGluPro-----AlaGl  138

Qy   702 -------------AGCCAACAAGAAATATAGCCCAGTGGGCAACTT 734
Qy       180 TCAAGCTCCTAACCAATCACCACCCCCTACT------TC  212

Db   562 GlnGlyGlyGlnProGlyTyrTyrProThrSerProGlnGlnIle-------GlyGlnGly 578
Db       138 nGluAlaProLysSerValAlaProThrProGlnProLeuAlaProProValProSe  158

Qy   735 GAAAAGGAGGAGATGCAGCAGATGGGCTGGCTGATGAACAATCACATACATACCCAGAAGCTCAATCCTTT 794
Qy       213 ACACCAGAAAGTCAGATGCCCCAAAGTCGCCA---------------  246

Db       579 Gln-------  579
Db       158 rLeuGlnSerProSerLysProAlaProThrProProAlaLysAlaAlaPr  178

Qy   795 CAGTGAAGAGTTGATTACGAGCTGCCATGTCCACCCGCCTACACCAAAGGAGATGAGGG 854
Qy       247 ----CCCCGCTGCCAGAAGGACATGGACATGGAAAGTCAGAGAAGCCCCTGAGGT  302

Db       580 -----------GlnProGlyLnLeuGlnGlnLnProThrGlyLnGlyGlnGlnGly 593
Db       178 oAlaProArgLeuAlaGlyProProGlyArgThrAlaSerProAsnLysThrAlaValPr  198

Qy   855 CTATGGCGCCCCAA-----------AGAAGGAACCAAAAC 884
Qy       303 TTCTCATCATCAAAAGAAAGAGGTGTCCAAAACGTGTCAGCAAGACTTACGAGAGAGG  362

Db   594 GlnGlnProGlyLyGlnGlyGlnGlnGlnLnProGlyLnGlyGlnGlnGlnGln 613
Db       198 oAlaProAlaLysProLysValAsnArgProGluIleValSerLeuLysAspAsnArgGl  218

Qy   885 TGCTGAAAGGGGCAAAGCGCTGCTGAGGAGCACATCTACAGGGAAATGATGAACATGTG--- 941
Qy       363 A--------------------  405

Db   614 ProGlyGlnGlyln------------GlnProGlyGlnGlyGlnProGlyTyrTyrPro 629
Db       218 yGlnGluGlyValAlaI1  232

Qy   942 -----CTTCATTATCTGCACAATGGCTCGCCACAGACAAGGATGGCAAGATCCAGGT 992
Qy       406 -GGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCAGCAGCCACGG  464

Db   630 ThrSerLeuGlnGlnSerGlyLnGlyGlnGlnProGlyGlnGlnProGlnGlnProGly 648
Db       232 eAlaAlaProGluProProLysProLysProLysValGluLeuArg--------Pr  247

Qy       465 CTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAG 524

Db       247 oLysProProArgProGluAspGluAsnLeuProGluLeuLeu----------  262

Qy       525 AGTGATGGAGGCAGGAGGAGCCCACATGGAGGAGGATCGAGAGCGTAGACACAGACAGCAGCGG 584

Db       263 --------GluPheProProLeuSerArgGlyHLysGlyValAspGlyAspAsnAspAl  279

Qy       585 CTATGGAGGAGAG-----------GCTGAGGAGGAGGCCCGAGCCGATGGAGTGGTGCTGT 638

Db       279 aAspAspGlyAspLeuSerThrGlyLysProLysPro-----------  292

Qy       639 GGTGAGGATCAAGCGGCCCCTTGCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTG 698

Db       293 ---LysLeuLysArgProThrProProArgLeu-------------  302

Qy       699 CAAAGCCCACAACAGAGAAATATAGCCCAGTGGGCAACTTGAAAGGAGATGAGCAGTGGGC 758

Db       303 ----------------  303

Qy       759 TGATGAACACATACAATCCCAGAAGCTCAAT  789

Db       327 gArgProLysMetAspAspAspAspAspLeuAspIleAsp-  342

Qy       790 ----CCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAAGG 845

Db       342 YAspAsnGlyProLysProThrLeuValSerLeuSerIleAlaArgProLysProLys 362

Qy       846 AGATGAGGGC-----TATGGCGCCCC  867

Db       868 -----AAAGAAGGAACCAAAAACTGCTGAAAGGGGCCAAGCGT 903

Qy       362 sSerLeuAlaAlaLysProSerThrProThrValAlaLysValLysLys 378
```

RESULT 14

S74544

translation initiation factor IF-2 - Synechocystis sp. (strain PCC 6803)

N:Alternate names: hypothetical protein slr0744

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001

C:Accession: S74544

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74544

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1001 <KAN>

A;Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAA16696.1; PIDN:g1651768

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C;Genetics:

A;Gene: infB

C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu homo

C;Keywords: GTP binding; nucleotide binding; P-loop

F;496-609/Domain: translation elongation factor Tu homology <ETU>

F;502-509/Region: nucleotide-binding motif A (P-loop)

F;606-609/Region: GTP-binding NKXD motif

F;642-644/Region: GTP-binding SAK/L motif

F;508,509,529,606,607,609,642/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #

Alignment Scores:	
Pred. No.:	0.672
Score:	116.50
Percent Similarity:	33.61%
Best Local Similarity:	22.69%
Query Match:	5.66%
DB:	

Length:	1001
Matches:	81
Conservative:	39
Mismatches:	111
Indels:	126
Gaps:	18

RESULT 15
T46337
hypothetical protein DKFZp43402413.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C;Accession: T46337
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23037
A;Accession: T46337
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-992 <AAA>
A;Cross-references: EMBL:AL137265
A;Experimental source: adult testis; clone DKFZp43402413
C;Genetics:
A;Note: DKFZp43402413.1

Alignment Scores:
Pred. No.:   0.732
Score: 116.00
Percent Similarity: 34.79%
Best Local Similarity: 20.74%
Query Match: 5.64%
DB: 2

Length: 992
Matches: 90
Conservative: 61
Mismatches: 147
Indels: 136
Gaps: 17

US-10-644-659A-1 (1-1146) x T46337 (1-992)

Qy   127 AGGCAGGCCAGGAGCTCACAGGCTGGCTG----------CCGGGAGGGACCCAGGAC 174
              |||||||||||||||| |||| |||||||
Db   212 ArgGlnAlaGlnGlnProLeuGlyIleGluAspLysAspAspSerGlnSerGlnAsp 231

Qy   175 TCACCTCAAGCTCCTAAACCA----------------ATCACCACCCTACTTCACAC 216
                        ::: :: :::
Db   232 GluLeuGlnSerLysGlySerGlyLeuGluArgLeuGluSerProLeuProHis 251

Qy   217 ---CAGAAAGTCAGAGTGCCCAAAGTCG----------------CCACCCGCCTG 255
              :::::|||||||||| |||::||||
Db   252 GluGluArgAlaGlnSerProArgSerProArgSerLeuAlaThrGluGluProGlnGly 271

Qy   256 CCAGAAGGACACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTCT--------- 306
              |||||||| ::: ::: ::||||||:::::
Db   272 ProGluGlyGlnProGluTrpLysGluAlaGluLeuGlyLeuGluLeuGlyLeuGluAspSerAlaAlaSer 291

Qy   306 --------- 306
Db   292 LeuSerLeuGlnLeuSerLeuArgGluGlnAlaProSerProProAlaAlaCysGlu 311

Qy   307 ------CACATCAAAAGAAAGAGGTG---------- 327
               :: ::: :|||||:::
Db   312 LysGlyLysGluGlnHisSerGlnAlaGluGluLeuGlyProGlyGlnGluGluAlaGlu 331

Qy   328 --------TCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGAGGGGACGTG----- 372
                     :::::
Db   332 AspProGluGluLysValAlaValSerProThrProProValArgSer 351

Qy   373 ------------------------------AGCCACCTCAGCCACCAGGTACGAG 396
                                           ::: ||| |||
Db   352 ThrGluProValAlaProProGluGlnLeuSerGluAlaAlaLeuLysAlaMetGluGlu 371

Qy   397 AGGGATGCTGGTGTGCTTGAACCTGGGCAGCCA---------- 429
              |||||| |||||||
Db   372 AlaValAlaGlnValLeuGluGlnAspGlnArgHisLeuLeuGluSerLysGlnGluGluLys 391

Qy   430 ------------------GAGAATGACATTGACAGAATCCTCCACAGCCAC 462
                                ::: ::: |||||| |||||
Db   392 MetGlnLeuArgGluLysCysGlnGluGluGluGlnGluIleLeuLeuArgLeuHis 411

Qy   463 GGCTCCCCAACGGCGGAGAGAAAATGTGCCAACCTGTGTCTGAGCTAACCAAGGGCTGG 522
              ::::||||||| ||| |||
Db   412 ------------------GlnGlnLysGluGlnSerLeuSerSerLeuSerLeuArgGluArgGluLeuGln 426

Qy   523 AGAGTGATGGAGCAGGAGGAGGAGGCCCATGGAGGAGGTGACAGC---------- 564
              ||| :::||||||::: |||||||
Db   427 LysAlaIleGluGluGluGluGluAlaArgMetArgGluGluSerGlnArgGlnLeuSerTrp 446

Qy   565 -----GTAGACACAGGAGCAGCGGCTATGGAGGACGGCTGAGGAGGAGGCCGAGCAG 618
              :::
Db   447 LeuArgAlaGlnValGlnSerSerThrGlnAlaAspGluAspGlnIleLeuArgAlaGluGln 466

Qy   619 GATGGAGTGCAGGTGGGTCAGGATCAAGCGCCCCCTTGGCCTCCCAGGTAAACAGA 678
              ::: :::::::: :::::::
Db   467 ------------GluAlaSerLeuGlnLysLeuArgGluGlnLeuLeuGlnLysGlnLysLysAla 483

Qy   679 TTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAA 738
              ||| :::|||||| |||:::|||
Db   484 GluArgAlaSerArgLeuGlnLysGlnLysAsnArgGlnMetLeu- 496

Qy   739 GGGAGATGGCAGCAGTGGCTGATGAACAATCACATCCAGAAGCTCCAATCCTTTCAGT 798
              :::|||| ::: ::||||
Db   497 ----------GluGlnLeuLysGlyGluIleGluAlaSerGluLys- 509

Qy   799 GAAGAGTTTGATTACGAGCTGCCATGTCCACCCGCCTACACAA--GGAGATGAGGGC 855
              |||||| ||| ||| :::::
Db   510 GluGlnAlaAlaLeuAlaAsnAlaAlaLysGluLysAlaLeuAsnLeuGlnLeuArgGluGlnLeu 529

Qy   856 TATGGCCGCCCCAAAGAAGACCAAAACTGCTGAAAGGGCCAAGCCTGCTGAG------- 909
              :::||| ::::|||
Db   530 GluGlyGluArgLysGluAlaValAlaAlaThrLeuGlyLysGluLysHisSerAlaGluLeuGlu 549

Qy   910 ------------------GAGCACATCACAGGAAAATGATGGACATGTGCTTCATTATC 951
                                ::::||||:::|
Db   550 ArgLeuCysSerSerLeuGluAlaLysHisArgGluValValSerSerLeuGlnLysLysLys 569

Qy   952 TGCCAATGCTGCCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTT 1011
              ||||| ::: :: :::::|||
Db   570 IleGlnLeuAlaGlnGlnLysLeuGlnLeuAlaGlnLysCysLeuGlyGlnValGlu 589

Qy   1012 GACAGATACGTTCGTATTTCGTATTCGGATAAAGTAGTGGGCATT-----------CTC 1053
              ||||| :::::
Db   590 HisArgValHisGlnLysSerTyrHisValAlaGlyTyrGluHisGluLeuSerSerLeu 609

Qy   1054 ATGCGTGCCAGGAAAACATGGACTGGTAGACTTTGAAGGAGAG 1095
              ::|| |||||||
Db   610 ----------GluValGluGlyGlu 620

Qy   610 LeuArgGluLysArgGln---------

Search completed: May 4, 2004, 08:55:57
Job time : 40 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:     May  4, 2004, 08:49:09 ; Search time 20 Seconds
                              (without alignments)
                              5916.336 Million cell updates/sec

Title:          US-10-644-659A-1
Perfect score:  2058
Sequence:       1 atggctccggcgaaagga..........tgattacgctactcaagtga 1146

Scoring table:  BLOSUM62
                Xgapop  10.0 , Xgapext    0.5
                Ygapop  10.0 , Ygapext    0.5
                Fgapop   6.0 , Fgapext    7.0
                Delop    6.0 , Delext     7.0

Searched:       389414 segs, 51625971 residues

Total number of hits satisfying chosen parameters:       778828

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10644659/runat_04052004_084903_28566/app_query.fasta_1.1287
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=Bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10644659 @CGN_1_1_27 @runat_04052004_084903_28566 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     Issued_Patents_AA:*
     1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
     2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
     3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
     4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
     5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
     6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	DB	Length	ID	Description
1	138.5	6.7	4	748	US-09-252-991A-18427	Sequence 18427, A
2	134.5	6.5	4	454	US-09-252-991A-30164	Sequence 30164, A
3	134	6.5	4	416	US-09-252-991A-30923	Sequence 30923, A
4	133.5	6.5	4	1225	US-09-252-991A-25018	Sequence 25018, A
5	130.5	6.3	4	442	US-09-252-991A-31848	Sequence 31848, A
6	130.5	6.3	4	1739	US-09-976-594-76	Sequence 76, Appl
7	130	6.3	4	586	US-09-252-991A-28601	Sequence 28601, A
8	127.5	6.2	4	499	US-09-252-991A-18439	Sequence 18439, A
9	127	6.2	4	638	US-09-252-991A-27068	Sequence 27068, A
10	126.5	6.1	4	609	US-09-252-991A-23771	Sequence 23771, A
11	125	6.1	4	741	US-09-252-991A-27062	Sequence 27062, A
12	124.5	6.0	4	256	US-09-252-991A-25404	Sequence 25404, A
13	124.5	6.0	4	482	US-09-252-991A-17580	Sequence 17580, A
14	124	6.0	4	1113	US-09-252-991A-29215	Sequence 29215, A
15	121.5	5.9	4	437	US-09-252-991A-26942	Sequence 26942, A
16	121	5.9	4	422	US-09-252-991A-27632	Sequence 27632, A
17	121	5.9	4	1182	US-09-287-354-6	Sequence 6, Appli
18	120.5	5.8	4	728	US-09-252-991A-23613	Sequence 23613, A
19	120	5.8	4	507	US-09-252-991A-28026	Sequence 28026, A
20	119.5	5.8	4	479	US-09-252-991A-32884	Sequence 32884, A
21	119.5	5.8	4	507	US-09-252-991A-23598	Sequence 23598, A
22	118.5	5.8	4	508	US-09-252-991A-25549	Sequence 25549, A
23	118.5	5.8	4	529	US-09-252-991A-18245	Sequence 18245, A
24	118	5.7	4	407	US-09-252-991A-31517	Sequence 31517, A
25	118	5.7	4	863	US-09-252-991A-26099	Sequence 26099, A
26	117.5	5.7	4	429	US-09-252-991A-33116	Sequence 33116, A
27	117.5	5.7	4	52	US-09-198-452A-52	Sequence 52, Appl
28	117.5	5.7	4	528	US-09-252-991A-19913	Sequence 19913, A
29	117.5	5.7	4	822	US-09-252-991A-22479	Sequence 22479, A
30	117	5.7	4	373	US-09-252-991A-29008	Sequence 29008, A
31	117	5.7	4	379	US-09-252-991A-28428	Sequence 28428, A
32	117	5.7	4	625	US-09-252-991A-19871	Sequence 19871, A
33	117	5.7	4	788	US-09-252-991A-28171	Sequence 28171, A
34	117	5.7	4	989	US-09-252-991A-17435	Sequence 17435, A
35	116.5	5.7	4	593	US-09-252-991A-31061	Sequence 31061, A
36	116.5	5.7	4	693	US-09-252-991A-24059	Sequence 24059, A
37	116	5.6	4	411	US-09-252-991A-25491	Sequence 25491, A
38	116	5.6	4	414	US-09-252-991A-31110	Sequence 31110, A
39	116	5.6	4	542	US-09-252-991A-21753	Sequence 21753, A
40	116	5.6	4	822	US-09-252-991A-21920	Sequence 21920, A
41	115.5	5.6	4	277	US-09-252-991A-25033	Sequence 25033, A
42	115.5	5.6	4	300	US-09-252-991A-27253	Sequence 27253, A
43	115.5	5.6	4	449	US-09-252-991A-23908	Sequence 23908, A
44	115.5	5.6	4	564	US-09-252-991A-28765	Sequence 28765, A
45	115	5.6	4	273	US-09-252-991A-23651	Sequence 23651, A

ALIGNMENTS

```
RESULT 1
US-09-252-991A-18427
; Sequence 18427, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J.  Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18427
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18427

Alignment Scores:
Pred. No.:                              3.53e-05
Score:                                  138.50          Length:         748
Percent Similarity:                     30.92%          Matches:        97
Best Local Similarity:                  24.19%          Conservative:   27
Query Match:                            6.73%           Mismatches:     111
DB:                                     4               Indels:         166
                                                        Gaps:           21

US-10-644-659a-1 (1-1146) x US-09-252-991A-18427 (1-748)

Qy   23 GCGGGAGGGCCCAGCCAAGAGCGCCCTCCGGA-------AGATACGCACAGCCACCC 73
            ||:|||||   |||:|||:|||::||||          |||  ||||||||||
Db   88 AlaGlyThrAlaArgProArgArgAsnArgGlySerArgProArgProAlaGlnPro--- 106
```

```
Db   397 gArgAsnLeuArgMetGlyArgProAlaArgArgArgArgAlaValAlaGlnGlyLeuProAlaHi 417
            :::||      |||         ||               ||||||
Qy   938 T 938
Db   417 s 417
```

RESULT 2
US-09-252-991A-30164
; Sequence 30164, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30164
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30164

Alignment Scores:
Pred. No.:   7.04e-05   Length: 454
Score: 134.50   Matches: 85
Percent Similarity: 35.47%   Conservative: 48
Best Local Similarity: 22.67%   Mismatches: 147
Query Match: 6.54%   Indels: 96
DB: 4   Gaps: 15

US-10-644-659A-1 (1-1146) x US-09-252-991A-30164 (1-454)
```

```
Qy    11 GCGAAAAGGAAAGCGGGAGGGGGAGGCCCAGCCAGGAGGGCCCTCCGGAAGAGATACGCA 64
Db     5 AlaProArgProAlaGly----------ProAlaProAlaGlyArgGlyAlaArgGlyAlaProGly 20

Qy    65 ------CAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCCGCCGAGGTGGCGCGA---ATG 115
Db    21 GluGlySerProGlyTrpSerProAlaProGlyGlu------ThrAlaGlyArgArgCysThr 38

Qy   116 AGAACAGGATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTCGGTCTGCCGGGAGGGGACCCCAG-GAC 174
Db    39 ArgArgAlaAlaGly----------AlaAlaCysArgArgAlaProAlaProGlySer 52

Qy   175 TCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGGAAAGCTCACAGAGTGCC 234
Db    53 SerArgLysSerProSerAlaLeuAlaProProAlaThrGlySerThrArgThrGlySerGlnPhe 72

Qy   235 CCA---AAGTCGCCACCCGCCTGCCAGAAGGA 264
Db    73 ProCysSerSerProProGlyLeuAlaGlyLyArgArgAlaAlaSerSerProTrpAla 92

Qy   265 ---------------------------------CATGGA 270
Db    93 AlaArgSerArgArgSerTrpArgLeuGlnHisGlyThrGlnGlnAlaGlnLeuArgVal 112

Qy   271 GATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTGTCC 330
Db   113 AspGlyGluGlnGlnGlnGlyIleValGlyValLeuHisGluAspAlaGlnValGlnHisGln 132

Qy   331 AAAACGGTGGTCAGCAAGACTTACGAGAGAGGAGGGGACGTGAGCCACCTCAGCCACCAGG 390
Db   133 GlnGlnAlaGlyHisGlnAlaLeuAlaLeuAspArgHisGlyGlyAspGluArgHisGlyGlyValGlnValGlnHisValGlnArgArg 152

Qy   391 TACGAG---AGGGATGCTGGTCGTCTTGAACCTCGGGCAGCCAGAGAATGAC------------ 438
```

```
Qy    74 TGGTCATCAGCTTGGCCCGAGGTTGGC-----AGCAGTGGGCGAATGAGAACAGCATCA 127
Db   107 ----------------ValGlyGlnArgSerProGlySerArgArgThrArgArgPro 119

Qy   128 GGCAGGCCCAGGAGCCTACAGGCTGGCTGCCTGCCGGAGGGACCCAGGACTCACCTCAAGCTC 187
Db   120 HisArgProArgHis---SerGlyGlnCysArgArgPheAlaProGlyPheProGln 135

Qy   188 CTAAACCAATCACACCCCTACTTCAC------------ACC 217
Db   136 ---ArgProGlyHisProGlyGluHisGlnLnHisGlnArgArgGlySerSerProThr 154

Qy   218 AGAAAAGCTGCAGAGTGCCCAAAGTCGCCACCCGCCTGCCAGA----- 260
Db   155 GlnGlnProGlnArgProGlyArgArgHis-ProGlnProArgHisGlnLeuHisArgHi 174

Qy   261 ---AGGACATGGAGATGGACAAAGCTC--------------- 284
Db   174 sProArgArgArgTrpArgTyrArgProLeuGlyAlaArgLeuGlnArgProArgIleGlyAs 194

Qy   285 ----------------------------AGAGA 289
Db   194 pAlaThrValArgArgHisAlaProValHisArgHisGlyHisGlyGluLeuProGlnAr 214

Qy   290 AAGCCCCTGAGGTTTCTCACATCAAAAGAA--------- 320
Db   214 gSerValAspGlyArgAlaHisArgArgAspProArgProGlyLeuGlyLeuAlaValAlaValAlaArgAr 234

Qy   321 ---------------AGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGAGGGG 367
Db   234 gArgArgHisArgArgGlyArgAspGlnArgGlyAlaGluGluAlaPheArg--ArgArgAs 253

Qy   368 ACGTGAGCCACCTCAGCCACCAGGTACGAGAGGGATGCTGGTCGTCTTGAACCTCGGGCAGC 427
Db   253 pProGlnProProAlaProArgLeuArg------------LeuLe 264

Qy   428 CAGAGAATGACATTGACAGAATCCTCCA----------------- 455
Db   264 uArgGln------------ProProValGlyProProGlyGlnArgArgLeuAlaAlaAspAr 279

Qy   456 -CAGCCACGGCTCCCCACGCGCGAGGAG---- 482
Db   279 gGlnProGlnLeuProAlaGlnProPheArgGlnArgHisArgGluProAlaLeuAlaProProGluAspArgProTr 299

Qy   483 -AAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGCGGTGGAGAGTGATGGAGACAGGAGG 541
Db   299 ---------------------------HisGlnArgGlyAlaAlaLe 312

Qy   542 AGCCCACATGGAGGAGTGACAGCCTAGA--------CACAGAGGACAGCGCGGCTATGGAG 592
Db   312 uAlaGly----------GlnArgArgSerGlyLeuHisProGlyArgProArgLeuTrpArg 328

Qy   593 GAGAGGCTGAGGAGGAGGGCCGAGCAGCAGGATGGAGTGGAGTGCGTGTGGTGTCAGGATCAAGC 652
Db   328 gProGly-------AlaAspProGlyAlaArgLeuuArgHisProAlaGlyArgArgGlnValProGlnAl 347

Qy   653 GCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGAAAGCCCAACAGA 712
Db   347 aProAlaArgGluGluLeuGlnArgAlaGlnLeuArgArgAlaAlaAlaLeuGlnLnArgProValAs 367

Qy   713 A-------ATATAGCCCAGTGGCAACTTGAAAGGGAGGATGGCCAGCAGTGGG 757
Db   367 pProProAspGlnArgGluLeuProPheArgGlnArgHisArgGluGluGlnArgProValVal-- 386

Qy   758 CTGATGAACAACATACATCCAATCCTTCAGTGAAGAGTTTGATTACGAGC 817
Db   387 ---------------LeuHisGlnGlyPro- 391

Qy   818 TGGCCATGTCCACCCCGCCTACACAAGGAGGATGAGGGCTATGGCGCCCCAAAGAAGGAA 877
Db   392 -------------------------ProProLeuAlaGlnAr 397

Qy   878 CCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCATCTACAGAGAGACAATGATGGACA 937
```

```
Db    8  CGGGCGAAAAGGAAAGCGGGAGGGCCCAGCCAGCAAGAGCGCCCTCCGGAGATACGCA---   64
Qy   24  ArgAlaArgArgThrGlyAlaGlyLeuAlaArgArgProGlyArgGlyAla              43

Qy   65  ------CAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGCGAATGAGA     118
Db   44  ThrCysArgLeuProGlyGlnProArgTyrProArgThrGly------ProGlyArg       60

Qy  119  ACAGCATCAGGCAGCCCCAGGAGCCTACAGGCTGGCTGCCCGGGAGGGACCCAGGACTAC   178
Db   61  ---GlyArgProArgSerHisArgValGlyThrGlyValGlyGluArgGlnArgArgGly    77

Qy  179  CTCAAGCTCCTAAACCAATCACACCCCTACTTCACCACCAGAAAGTCAGA----------   229
Db   78  LeuArg-------------ArgProValLeuValArgArgValArgGlyArgAlaGlyGly    92

Qy  230  ------GTGCCCAAAGTGCGCCACCCGCCTGCCAGAAGGACATGGACATGGACAAAGCT    283
Db   93  GlyAlaValSerProGluGlnTyrProGly-------ArgAspValGlyLeuProGlnArg   110

Qy  284  CAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTGTCCAAAAC---------   335
Db  111  AlaLeuArgAlaLeuArgLeuGlyCysGlyValGluAspArgArgHisPro-GlnLeuProAl  130

Qy  336  ------GGTGGTCAGCAAGACTTACAGAGAGGAGGGGACGTGAGCCACTCAGCCACA     388
Db  130  aGlyAlaAlaGlyGlyAlaAspLeuArgAlaAlaArg---ArgHisProGlyProAl     149

Qy  389  GGTA-------CAGAGAGGGATGCTGGTGTGTTGAACCTGGCAGCCAGAGAATGACATTG   442
Db  149  aLeuLeuAlaArgAlaGly-AlaAlaMet-------GlyGlnProGlnArg-------    163

Qy  443  ACAGAATCCTCCACAGCCACGGCTCCCCAACCGGAGGAGAAAATGTGCCAACCTCGTG-   501
Db  164  ------------ArgProGlyAspArgArgArgArgCysAlaGlyArgCysArgArgValP  178

Qy  502  ------TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAGGAGCCCACAT       550
Db  178  roGlyArgArgGlyAspLeu-ProArg-------ArgAlaAlaPro------          190

Qy  551  GGAGGAGTGAGCAGCGTAGCACAGAGGACAGCGGCTATGGAGCAGGAGGC            610
Db  191  ------AlaAlaAlaProArgGlnArgProAlaAla---------AlaGly           202

Qy  611  CCGAGCAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCTCCCCAGG  670
Db  203  ProArgHisLeu-------------SerGlyProGlyAlaGlyThrProGlyArg        216

Qy  671  TAAACAGATTACAGAGAAACTCAACTGCAAAGCCAACAGAAAATAGCCCAGTGGGCA     730
Db  217  ------AlaValLeuAlaAlaGlyAlaGlyAlaThrLeu                        227

Qy  731  ACTTGAAAGGGAGGATGGGCAGTGGGCTGATGAACACATCAAT-----CCCAGAGAGG   784
Db  228  ------GlyHisGlyValValAlaAlaCysLeuAsnThrGluGlyGlyCysAlaArgArg   244

Qy  785  TCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGCCATGTCCACCCGCCTACCACAAAG  844
Db  245  ArgAlaAlaThrThrPheArgTrpArgThrGlyTrpThrCysProProAlaSerArgPro   264

Qy  845  ------GAGATGAGGGTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCCAAGC  901
Db  265  GlyArgPheArgCysArgArgSerProSerArg---------------ProGlyProAla   279

Qy  902  GTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATGG  961
Db  280  CysArgArgAlaThrThrAlaTrpLysProLeuProAlaSerThrThrAlaPro-AsnGl   299

Qy  962  CTCGCCACAGACCAGATGG----------CAAGATCCAGG                      991
Db  299  y------TyrCysArgTrpProAlaAlaArgArgSerArgProCysGlnGlyCysGl     317
```

```
Db  153  GlyGluAlaAlaAspGlnArgGlnAlaGluLeuGlyGluGluHisGlyAspGlnArg      172
Qy  439  ------ATTGACAGAATCCTCCAGCCACGGCTCCCCAAGCGGAGGAGAAAATGTGCC      492

Db  173  ArgGlyGluGlnArgGlyAlaGlyHisGlnHisLeuArgValGluLeuArgGluCysHis   192
Qy  493  AACCTGGTCTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAGCCCACATGG     552

Db  193  ProLeuAlaAlaGluLeu------------ArgGluGluThrArgAlaHis           204
Qy  553  AGAGTGACAGCGTAGACACAGAGACAGCGGCTATGGAGAGAGGCTGAGGAGGAGGCCC    612

Db  205  ArgProThrGlyIleAlaProGluGlnIleArgGlnAlaAspGlnArgGluGlyAlaAlaGly  224
Qy  613  GAGGAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCTCCCAGGTA   672

Db  225  GluGlnGluGlnGlyAspGln-------ArgLeuArgGlnGlnGlnProGlyGlyAla     240
Qy  673  AAC---------------AGATTTACAGAGAAACTC                          693

Db  241  GlyLeuAlaAlaGlnAlaGlyValGluGlyValGlyCysArgGlyAlaThrLeuValVal   260
Qy  694  AACTGCAAAGCCAACAGAAAATAGCCCAGTGGCAACTTGAAAGGGAGGATGGCAGCAG    753

Db  261  AspAspGluAlaGlyArgPheHisGlyProGluGlyAsnGlyGlyAspGluAlaAspGlu   280
Qy  754  TGGGCTGATGAACACATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTAC  813

Db  281  GlnAlaGluGlnHisPheLeuAla---------AspGlnGlnGlyAspVal           294
Qy  814  GAGCTGCCATGTCCACCCGCCTACACAAAGGAGGGCTATGGCCGCCCCAAAGAA        873

Db  295  GlnArgAlaGluGlyAlaArgGlnAlaAlaGlyAspAspArgValGlyGluGlnArgGln   314
Qy  874  GGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATG  933

Db  315  ArgThrGlyHisAlaValAlaAlaHisGlnAlaAlaAlaAlaGlnGlyLeu           329
Qy  934  GACATGTGCTTCATTATCTGCACAATGGCTCGCCACAGACCAGAT                978

Db  330  ------------------LeuSerGlnHisArgGlnAsp                       336
```

RESULT 3
US-09-252-991A-30923
; Sequence 30923, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30923
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30923

Alignment Scores:

Pred. No.:	7.58e-05	Length:	416
Score:	134.00	Matches:	111
Percent Similarity:	34.87%	Conservative:	40
Best Local Similarity:	25.64%	Mismatches:	140
Query Match:	6.51%	Indels:	142
DB:		Gaps:	27

US-10-644-659A-1 (1-1146) x US-09-252-991A-30923 (1-416)

Qy 149 lyArgArgGluProGlyThrAspArgCysGlyAlaArgValArgAlaIleAlaArgGlnTrpG 169
Db   :::!!:!!:!!!!!!   :::!!!!!

Qy 342 CAGCAAGACTTACGAGAGAGGGGACGTGAGCCACTCAGCCACGGTACGAGAGGGA 401
Db 317 y------------------------GlyAlaGlyGlySe 322

Qy 169 lnArgArgProAlaArgArgProGlyAlaProAlaAlaLeuAla 183
Db   !!!!!   !!!!!!

Qy 1052 TCATGCGTGCCAGGAA-------------------ACATG 1072
Db   ::!!!:!!!:

Qy 402 TGCTGGTGTGTCTGAACCTGGCAGCCAGAGAGATGACATTGACAGAGATCCTCCACAGCCA 461
Db 322 rAlaSerCysArgArgAlaMetProSerThrProGlyAlaArgProGlyThrTr 342

Qy   !!!!:::   ::: !!!!!

Qy 462 CGGCTCCCAACGCGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTG 521
Db 184 ----ValAlaProGlyProGlyLeuGlnArgPheGlyAlaAspGlnArgGlyAlaValAlaA 202

Qy 522 GAGAGTGAT-- 530
Db 202 laArgProGlu-----------------------SerGlnProArgAlaG 211

Qy 531 ------------------GGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACAC 572
Db 211 ly-AlaGlyGluArgGlnAlaProSerArgArgAspArgArgGlnHisGlnCysGlnArgArgAla 230

Qy 573 AGAGGACAGCGGCTATGGAGGAGGCTGAGGAGGCCCAGACGCAGGATGGAGTGCAGGT 632
Db 231 AlaGluArgLeuHisArgGlyArgProGlyTyrHisProAlaArgArgHisArgAlaHis 250

Qy 633 GGC-------------------TGTGGTCAGGATCAAGCGCCCCTTGCCTCCCCAGGTAAA 674
Db 251 ArgGlyAlaAla--------------ArgArgLeuGlyAlaValArgLeuGlyArgHisArgArg 268

Qy 675 CAGATT---------TACAGAGAAACTCAACTGCAAAGCCCAAGAGAAATAGCCCAGTGGG 728
Db 269 GlyAspGlnArgHisProGlnValGlyArgProArgProAlaArgProLeuArgArgGlnArg 288

Qy 729 CAACTTGAAAGGAGGAGATGGCAGCAGTGGCGTGATGACACATACAATCCCAGAAGCTCAA 788
Db 289 SerLeuArgArgTrpArgArgProAspProGluArgArgGlyGluTrpArgProAlaPhe 308

Qy 789 TCCTTTCAGTGAAGAGTTTGATTACGAGACTGCCATGTCCACCCG-------------- 833
Db 309   ::: !!!!!   !!!!!!   PheProGlnProGln 317

Qy 318 ------------ArgAlaValProArgThrGlyAlaAspHisProArgArgHisArgArgPro 333

Qy 834 --------------------CCTACACAAAGGAGGATGAGGGCTATGGCGCCC------ 866
Db 334 AlaTyrArgAlaLeuArgGlnProValHisArgArgProPheAlaAlaProProGlyAla 353

Qy 867 --CAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGGCTGCTGAGGAG----CACAT 917
Db 354   ::: !!!!!   ::: !!!!!!   :::!!!!

Qy 918 CTACAGGGAAATG-------------------ATGGACATGTGCTTCATTATC----TGCAC 956
Db 354 GlyLeuGlnArgArgArgGly-AlaGlyArgGlnArgArgArgArgValGlnLeuArgHisLe 373

Qy 957 AATG-------------------GCTCGGCCACAGCAGAGCTGCAAGGATCCAGGTTACTTT 998
Db 373 uTyrProSerLeuGlyValLeuGlyAlaAsnLeuProValAlaValProGlyAlaAlaTh 393

Qy 999 TGGAGAT 1005
Db 393 rLeuSerGlyLeuHisProAlaHisHisGlnArgArgGlyArgLeu--------Ph 410

Db 410 eAlaAsp 412

Qy 992 TTACTTTTGGAGATCTCTTGACAGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTC 1051
Db   !!!!!!   :::!!!

Qy 317 y-------------------------GlyAlaGlyGlySe 322

Qy 1052 TCATGCGTGCCAGGAA-------------------ACATG 1072
Db   ::!!!:!!!:

Qy 322 rAlaSerCysArgArgAlaMetProSerThrProGlyAlaArgProGlyThrTr 342

Qy 1073 GACTGGTAGACTTTGAAGGAGAGATGCTATGGCAAGGCCGAGA--------------- 1115
Db   !!!!!!!!   !!!!!!!

Qy 342 pCysGlyArg-----------------------SerAlaArgProArgSerSerProIleSerTh 356

Qy 1116 -----TGACCATGTGTGATTACGGTCACTCAA 1142
Db   :::!!!!!!:!!:

Qy 356 rAlaSerThrCysCysTrpTrpSerThrArg 366

Alignment Scores:

Pred. No.:	0.000143	Length:	1225
Score:	133.50	Matches:	100
Percent Similarity:	32.00%	Conservative:	36
Best Local Similarity:	23.53%	Mismatches:	156
Query Match:	6.49%	Indels:	133
DB:		Gaps:	18
			4

US-10-644-659A-1 (1-1146) x US-09-252-991A-25018 (1-1225)

Qy 12 CGAAAAGGAAAGCGGGGAGGGGCCCAGCCAAGAGCGCCCTCCGGAAGATACCGCCACGCCAC 71
Db   !!!!!!   :::

Qy 30 ArgValGlyGlyValArgGlyGlyProArgGlnHisGlnLnArgArgAspProAlaLeuArgHis 49

Qy 72 CCTGGTCATCAGCTTGGCCCCAGGTTGGCAGCAGTGGCGAATGAGACAAGCATCAGGCA 131
Db   ::: !!!!!

Qy 50 ProGlyArgArgThrGlyProGlyThrAlaGluHisSerProProGluSerHisPro-As 69

Qy 132 GGCCCAGGAGCCTACA---GGCTGGCTGCCGGGAGGGACCCCAGGAC--------- 174
Db   !!!!!!   :::

Qy 69 pArgLeuArgProAlaProGlyGlyValGlyProAlaGlyThrGlyThrGlyProGlyLeuLeuGl 89

Qy 175 ---TCACCTCCAAGC 185
Db   !!!!!!

Qy 89 nArgArgAlaGlyValAlaAlaAlaGlyLeuAlaGlyGlyValArgProGlyIleHisProGlnAr 109

Qy 186 TCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCCAAAGTCGCC 245
Db   ::: !!!!!

Qy 109 gArgArgArgAlaGlnProAlaThrArgHisGlyAlaGlyCysArgAlaAlaGlyGlyGl 129

Qy 246 ACCCCGCCTGCCAGAAGGACATGGAGATGGA--------------------CAAAG 281
Db   !!!!!!   !!!!!!

Qy 129 uAlaArgArgAspProGlyHisGlyGlyHisProGlnGlyArgHisArgProGlnGln-G 149

Qy 282 CTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGT 341

```
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31848
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31848

Alignment Scores:
Pred. No.:            0.000176        Length:          442
Score:                130.50          Matches:         109
Percent Similarity:   29.81%          Conservative:    29
Best Local Similarity: 23.54%         Mismatches:      121
Query Match:          6.34%           Indels:          204
DB:                                   Gaps:            24
```

US-10-644-659A-1 (1-1146) x US-09-252-991A-31848 (1-442)

```
Qy   15  AAAGGAAAGCGGGGAGGGCCCAGCCAAGAGAGCCCCTCCGGAAGATACGCACAGCCACCCT   74
             ::::|||||:::||||||||||||  |||  ||  |||
Db   62  ArgGlyArgArgGlyHisProProAspHisArgHisProProAlaArgArgLeuArgPro   81

Qy   75  GGT---CATCAGCTTGGCCCGAGGTTGGCAGCA-------------------------  104
             :::                    ||||||||||||
Db   82  GlyValProSerHisArgArgArgLeuProArgLeuAlaAlaArgArgArgArgArgArg  101

Qy  105  ----------------------------------------GTGGGGAATGAGAA  119
                                                       ||
Db  102  GlyGlnProAlaArgProGlySerGlyProAspAlaGlyAsnGlySerGlyGlyIleGlu  121

Qy  120  CAGCATCAGGCAGGCCAGGAGCCTACAGGCTGGCTGCCGGAGGACCCAGGACTCA--   177
             ||||  |||||  :::|||
Db  122  AspHis---AlaCysProGly-Pro         ProGlyGlyAlaGluAspProAr  137

Qy  178  -CCTCAAGCTCCTAAACCAATCACACCCCTACTTCACACCAGAAAGC              224
             |||||||:::||| |||  ||||||||  ||  |||  || |||
Db  137  gGlyHisProLeuProGlnHisProArgPro-SerArgProAlaLeuHis-ProProAla  156

Qy  225  TCAGAGTGCCCAAAGTGCCCACCCGCTGCCAGAAGGACATGGACGATGGACAAAGCTC   284
             ||||||  ||||||
Db  157  ThrLys------------ProProAspArgArgProGlyArg-------------  166

Qy  285  AGAGAAAGCCCCTGAGGTTTCTCACATCAAAAGAAAGAGGTGTCAAAACGGTGGTCAG   344
             ------------
Db  166  -----------------------------------------------------  166

Qy  345  CAAGACTTACGAGAGAGGAGGGGACTGAGCCCACCTCAGCCACCAGGTACGAGAGGGATGC   404
             ::|||||||||||
Db  167  ----AsnArgArgGluHisProArgArgArgProAlaValArgGln------  179

Qy  405  TGGTGTGCTTGAACCTGGCAGCCAGAGAGAATGACATTGACAGAAATCCTCCACAGCCACGG   464
             ||
Db  180  -------TrpArgArg------------------------------------  182

Qy  465  CTCCCCAACGCGGAGGAGAAAAATGTGCCAACCTGGTCTGTCTGAGCTAACCAAGGGCTGGAG   524
             :::||| :::::|||  ||||||||
Db  183  ---------AspAlaArgAspGlnSerCysHis---------ArgGlnHisGlyLeuAsp  196

Qy  525  AGTGATGGAGCAGGAGCAGCAGCAGCCCACATGAGGAGAGTGACAGCCTAGA--------   569
             |||||  :::
Db  197  LeuArgProAlaGlyAlaGlyAspHisProAlaLeuArgAspSerTyrProGly  216

Qy  570  ---------------------------------CACAGAGGACAGCAGGCGTATGAGGAGGAGGCTGA   602
                                               :::
Db  217  LeuArgAlaAspProArgHisGlnLeuHisArgGlyLeuAspGlnProArgArg-------  233

Qy  603  GGAGAGGGCCCGAGCAGCAGGAATGGCGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCC   662
             |||  |||||
Db  234  ------------------------------------------AlaAlaGlyProGly-  238
```

```
Qy  663  CTCCCAGGTAAACAGATTTACGAGAGAAACTCAACTGCAA-----------AGCCCAACAGAA   713
             :::   :::|||||||       ||:::
Db  239  ValProValAspArgArgHisArgGlyGlyGlyGlnLeuArgGlnCysGlnLeuProGluAsp   258

Qy  714  ATATAGCCCAGTGGCAACTTGAAAGGAGGAGATGGCAGCAGTGGCTGGCTGATGAACACATACA   773
             ::::||||
Db  259  -----ProProGlyArgLeuArgGlyArgArgGlyArgAlaGln------  268

Qy  774  ATCCCAGRAGCTCAATCCTTTCAGTGAGAGTTTGATTACGAGCTGCCATGTCCACCCG   833
             |||||||||||||  |||  |||||||  |||||
Db  269  ---ProGluAlaArg------------------------------------HisPro   274

Qy  834  --------------------------CCTACACAAAGGAGATGAGGGGCTATGGCCG   863
             |||  |||||
Db  275  ArgGlnGlnProAlaAspValLeuArgAspArgProGlyGlnArgAlaLeuGlyGlnArgProGln   294

Qy  864  CCCCAAAGAAGGAACCAAAAACTGCTGAAAGGCCAAGCCGTGCTGAGGAGCACATCTACAG   923
             |||  |||   |||  ||||||  ||:::
Db  295  ProArgAspArgProAlaAspProAlaLeuArgAspAspProArgLeuCysGlyGlySerSerLeuGln   314

Qy  924  GGAAATGATGACACATGTGCTTCATTATCTGCACAATGGCTCG---CCACACGAGGATGG   980
             ||||||
Db  315  AlaValProGlyGluHisGlyArgLeuHisArgSerArgValProValGlnArgGln   334

Qy  981  CAAGAT--------------------CCAGGTGACTTTTGGAGATCTCTTTGACAGATACGTTCGTAT   1028
             |||||||  :::
Db  335  AlaAspHisProArgProGly----------ArgProLeu------LeuArgGlnAla   349

Qy  1029 TTCAGATAAAGTAGTGGCATTCTCATGCGTCGTCCGAGGAAACATGGACTGGT-------   1079
             ||||||   :::
Db  350  AlaArgAlaAlaAspGlyLeuArgHisLeuLeuHisGlnProCysArgGlyArgProGly   369

Qy  1079 ------------------------------------------------------  1079
             
Db  370  ArgHisGlyHisAlaProAspProAlaGlyArgArgArgPproGlnLeuHisHisGlyHis   389

Qy  1080 ---AGACTTTGAAGGAGGAGAGTGCTATG-----------------GCAAGGCCGAGATGACCCATGT   1124
             |||
Db  390  SerArgLeuArgArgArgArgAspAlaGlnLeuProAspHisLeuLeuProArgArgProLeu   409

Qy  1125 TGT 1127
        |||
Db  410  Cys 410
```

RESULT 6
US-09-976-594-76
; Sequence 76, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROI
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 76
; LENGTH: 1739
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 3125723CD1
US-09-976-594-76

```
Alignment Scores:
Pred. No.:            0.000341        Length:          1739
Score:                130.50          Matches:         63
Percent Similarity:   37.10%          Conservative:    42
```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28601
; LENGTH: 586
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28601

Alignment Scores:
Pred. No.:   0.000226   Length: 586
Score: 130.00   Matches: 82
Percent Similarity: 33.64%   Conservative: 29
Best Local Similarity: 24.85%   Mismatches: 132
Query Match: 6.32%   Indels: 87
DB: Gaps: 13

US-10-644-659A-1 (1-1146) x US-09-252-991A-28601 (1-586)

```
Qy      2 TGGCTCCGGGCGAAAAGAAAGACGGGGAGGGGCCCAAGAGCG------------CCC   49
                                                              :::
Db    128 TrpPheArgThrAlaAlaArgGlyGlyGlyAlaGlnProAspAlaGlyProProAspPro  147

Qy     50 TCCGGAAGATACGACAGCCACCCTGGTCATCAGCTTGGCCGAGGTTGGCAGCAGTGGG   109
          ||||||||                          :::|||
Db    148 ProArgLeuAlaArgLeuProProArgGlnAlaAlaAsnProSerArgArgProGly   167

Qy    110 CGAATACGAATG                          ---ACAGCATCAGGCAGGCCCAGAGCCTACAGGCTGGC   154
          |||:|:|||                              :::
Db    168 ArgLeuArgArgAspProHisProLeuAlaAlaAspArgProArgAlaLeuAlaAla   187

Qy    155 TGC------------------------------------------------CGGGAGGGA   166
          |||
Db    188 AlaProAlaGlnArgArgGlyGlyAlaAlaAspArgAlaAlargThrHisAspHisGluArg   207

Qy    167 CCCAGGACTCACCTCAAGCTCCTAAACCAATCA-----CACCCCTACTTCACACC   217
          :::|||                                              |||
Db    208 ProGlySerHisLeuArgLeuArgArgHisGlyProGlyHisProGlyMetHisAla   227

Qy    218 AGAAAGCTCAGAGTGCCCCAAAGTCGGCACCCC-----GCCTGCCAGAAGGACATGGAGGAGATG   274
          :::|||                                          |||
Db    228 AlaArgLeuCysLeuProGlyValArgArgProGlyValAlaGlyProLeuLeuValArg   247

Qy    275 GACAAAGCTCAGAGAACGCCCTGAGGTTTCTCACATCAAAGAAAGAGAGTGTCCAAAA   334
                |||  |||||||
Db    248 ArgArgArgGlnHisGlnProLeuLeu-ArgGlyGlyGlnAlaAlaThrGly---GlnAs   266

Qy    335 CGGTGGTCAGCAGACTTACGAGAGAGGAGGGGCACCTCAGCCCAGCAGGTACG   394
                |||  |||||
Db    266 pProGlyLeuArgArgLeuArg---ArgLeuLeuProProProGlyPro-------   282

Qy    395 AGAGGGATGCTGCTGTGTCGTTGAACCTGGGCAGCAGAATGACATTGACAGAATCCTCC   454
          :::
Db    283 ------------------------------------------------GlyProAl   285

Qy    455 ACAGCCACGGCTCCCCAACCGCGGAGAGAGAAAATGCCAACCTGGTGTCTGAGCTAACCA   514
          |||||||||||||||                    :::
Db    285 aLeuAlaProAlaArg------LeuAlaProAlaAsn-Arg-------LeuAlaIleArgArgLeuGlyL   302

Qy    515 AGGGCTGGAGAGTGATGGAGAGGAGGGCCCACATGGAGGAGCTAGACACAG   574
          |||                                   :::
Db    302 ysArgGluLysAlaLeuAlaProAlaSerProGluLysArgGlnPheAspIleThrAspT   322

Qy    575 AGGACAGCAGCGGCTATGGAGGAGAGGCTGAGGAGGAGCCCGAGCAGGATGGAGTGCAGGTGG   634
                ||||||          |||:|||
Db    322 rpAspSerSerPhe-ArgLysArgArgArgArgAla------------------------   333
```

```
Qy    109 GCGAATGAGAACAGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCCGGGGAGGACC   168
          :::|||||     ::|:|||
Db     17 SerAsnAlaSerHisSerAlaSerGluGlnAlaSerGlyGluAlaSerGlySerGlySerGln   36

Qy    169 CAGGACTCCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCACAC--CAGAAAGCT   225
          :::|||          |||          ::::::
Db     37 SerGluSerGluGlnGlySer--------AspProGlySerGlyHisGlySerGluSer   53

Qy    226 CAGAGTGCCCCAAAGTCGCCACCCGCTGCCAGAAGGACATGGAGATGGACAAAGCTCA   285
               |||||          :::|||
Db     54 AsnSerSerGluSerGluSerGlnSerGluSerGluSerGluSerAlaGlySer   73

Qy    286 GAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAGAGGGTGTCCAAAAACGTGGTCAGC   345
          :::          ::|||
Db     74 LysSerGlnProValLeuProGluAlaLysGluLysGluProAlaSerLysLys--------   90

Qy    346 AAGACTTACGAGAGAGGAGGGGACGTGACAGAGCGTGACGAGGAGGATGCT   405
               :::          :::
Db     91 --------GluArgIleAlaAspValLysLysMetTrpGluTyrProAspValTyr   107

Qy    406 GGTGTGCTTGAACCTGGGCAGCCAGGAGAATGACATTGACAGAATCCTC------   453
          ||||||          :::
Db    108 GlyValArgArgSerAsnArgSerArgArgGlnGluProSerArgPheAsnIleLysGluGlu   127

Qy    454 ------CACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTG   501
                             |||:::|||
Db    128 AlaSerGlySerGluSerGlySerProLysSerArgArg   140

Qy    502 TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGCAGGCCCACATGGAGGAGGAC   561
          |||||||||          :::
Db    141 --------GlyGlnArgGlnIleLeuLysLysGlyLeuLysTrpLysGlnLysPro   155

Qy    562 AGCGGTAGACACAGAGGACAGCGGCTATGGAGGAGAGGCTGAGGAGGAGCCCGAGCAGGAT   621
          |||          :::
Db    156 SerGluAspGlyGlnGlyLeuGlyThrSerArgAlaGluSerGlu------ProGluGln-   172

Qy    622 GGAGTGCAGGTGGCTGTGTCAGGATCAAGCGCGAGGATCAAACAGATTT   681
          ||||||          :::
Db    173 ----------LysLysValLysValAlaArgArgProValPro-------ArgArgThr   185

Qy    682 ACAGAGAAACTCAACTGCAAAGCCCAACAGAAAATAGCCCAGTGGGCAACTTGAAAGGG   741
          :::          |||||
Db    186 ValProLysProArgValLysLysGlyGlnProLys--------ThrGlnArgGlyLysArg   202

Qy    742 AGATGCAGCCAGTGGGCTGATGAA------------------GGGTAGCAGAAGACCAGCAAAAC   765
          :::          |||||||
Db    203 LysLysGlnAsnSerSerArgAspGluAspAspAspAspGluAlaProLysArgGlnThr   222

Qy    766 CACATACAATCCCAGAAGCTCAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCCATG   825
          :::          |||||:::|||
Db    223 ArgArgArgAlaAlaLysAsnValSerTyrLysGluAlaAspAspAspPheGluThrAspSer   242

Qy    826 TCCACCCGCCTACACAAAGGAGATGAGGGCTATGCCGCCCCAAAGAAGGAACCAAAACT   885
          |||          :::::|||
Db    243 AspAspLeuIleGluMetThrGlyGluGlyValAspGlyGlyValAspGlnGlnAsnSerGluThr   262

Qy    886 GCTGAAAGG   894
          |||
Db    263 IleGluLys   265
```

RESULT 7
US-09-252-991A-28601
; Sequence 28601, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```
Qy    635  CGTGTGGTCAGGATCAAGGCGCCCTTGCCCCTCCCAGGTAAACAGATTTACAGAGAAACTCA  694
                 ||||||||    ||||||||
Db    334  ---TrpLysAlaSerSerAla-----------------------------ArgSerArgGlyThr  344

Qy    695  ACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGT  754
                 |||||                                           :::
Db    345  AlaTrpLysPro------------------------------------CysSerAspProAlaHis  354

Qy    755  GGGCTGATGAACACATACAATCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACG  814
                 ||||||  :::       |||||||    :::
Db    355  GlyValGlnHisAlaArgValProArgCysPheValThrIleProAlaProAlaAlaGlu  374

Qy    815  AGTCGCCATGTCCACCGCTACCACAAAGGAGATGAGGGCTATGGCCGCCCAAAGAAG  874
                 :::  |||                                      :::
Db    375  AlaGlyProAlaPheProThrLeuGlyValValArgAlaArgAlaAlaCys----------Arg  391

Qy    875  GAACCAAAACTGCTGAAAGGGCCA  898
                 |||::: ||||||
Db    392  ArgProArgArgLeuAlaGlyPro  399
```

RESULT 8
US-09-252-991A-18439
; Sequence 18439, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Marc J. Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 107196.136
;   CURRENT APPLICATION NUMBER: US/09/252,991A
;   CURRENT FILING DATE: 1999-02-18
;   PRIOR APPLICATION NUMBER: US 60/074,788
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: US 60/094,190
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
;   SEQ ID NO 18439
;   LENGTH: 499
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18439

Alignment Scores:
Pred. No.:   0.000374   Length: 499
Score: 127.50   Matches: 109
Percent Similarity: 30.10%   Conservative: 43
Best Local Similarity: 21.58%   Mismatches: 162
Query Match: 6.20%   Indels: 191
DB: 4   Gaps: 23

US-10-644-659A-1 (1-1146) x US-09-252-991A-18439 (1-499)

```
Qy    23   GCGGGAGGGCCAGCCAAGAGCGCCCTCCGGAAGATACGCACAGCCACCCTGGTCATCA  82
                 ||                          |||:::||||     :::
Db    3    AlaAlaThrThrProSerAlaProLys---ArgTyrSerArgProAspTrpProThr  21

Qy    83   GCTTGGCCCGAGGTTGGC------------------------------AGCAGTGGG---  109
                 |||||:::        ||||||
Db    22   ArgTrpArgSerAlaGlyArgSerProThrProThrTrpSerSerGlyValSer  41

Qy    109                                                              109
```

```
Qy    110  ----------CGAATGAGAACAGCATCAGGCAGGCCCAGGAGCCTA-----------------  145
                      :::   |||||  ||||||    :::
Db    42   AlaArgArgTrpProArgProThrArgArgGlySerThrValAlaThrLeuProAlaThr  61

Qy    146  ---------------CAGGCTGGCTGCCGG----------GAGGGA  166
                            :::       |||||||
Db    62   ProIleThrArgLeuSerAlaSerProLeuArgProAspSerLeuProSerHisAlaSer  81

Qy    167  CCCCAGGACTCCACCTCAAGCTCCTAAACCAATCACCACCCCTACTTCACCAGAAAGCTC  226
                 |||||||||                            :::
```

```
Db   432  --ArgHisArgArgProArgArg------------GlnLeuGlyAspArgArgG  445
Qy  1022  TTCGTATTCAGATAAAGTAGTAGTGGCGATTCTCATGCGTGCCAGGAAACATGGACTGGTAG  1081
               |||                       |||    |||  |::
Db   445  lnArg---------------ProGlyAlaAlaArgArgGlyArgArgAlaGlyLeuProG  460
Qy  1082  ACTTTGAAGGA  1092
               :::|||
Db   460  lyGlyAspGly  463

RESULT 9
US-09-252-991A-27068
;  Sequence 27068, Application US/09252991A
;  Patent No. 6551795
;  GENERAL INFORMATION:
;    APPLICANT:  Marc J. Rubenfield et al.
;    TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;    TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;    FILE REFERENCE:  107196.136
;    CURRENT APPLICATION NUMBER:  US/09/252,991A
;    CURRENT FILING DATE:  1999-02-18
;    PRIOR APPLICATION NUMBER:  US 60/074,788
;    PRIOR FILING DATE:  1998-02-18
;    PRIOR APPLICATION NUMBER:  US 60/094,190
;    PRIOR FILING DATE:  1998-07-27
;    NUMBER OF SEQ ID NOS:  33142
;  SEQ ID NO 27068
;    LENGTH:  638
;    TYPE:  PRT
;    ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-27068

Alignment Scores:
Pred. No.:              0.000474          Length:          638
Score:                  127.00            Matches:         79
Percent Similarity:     33.23%            Conservative:    27
Best Local Similarity:  24.76%            Mismatches:      123
Query Match:            6.17%             Indels:          90
DB:                     4                 Gaps:            13

US-10-644-659A-1 (1-1146) x US-09-252-991A-27068 (1-638)

Qy    11  GCGAAAAGAAAGCGGGAGGGAGGGCCCAG----------------------CCAAGA  43
               |||||   |||||
Db    53  AlaLeuArgArgThrArgAlaGlnHisProProValProArgArgGlnArgArgArgArg  72
Qy    44  GCGCCCTCCGGAAGATAACGGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGC  103
               ||||||::::|||||             |::|||
Db    73  AlaProAlaGlyArgAlaAlaSerPro--------------------------------  81
Qy   104  AGTGGGGCGAATGAGAACAGACAGCATCAGGCAGCCCAGGAGCCTACAGGCTGGCTGCGGGAG  163
               |||||:::||||||
Db    82  ----------------GlyLysProArgArgGlyArgGlyLeuAlaGlyAlaThrAla  93
Qy   164  GGACCCAGGACTCACCTCAAGCTCTAAACCAATCACACCCTACTTCACACCAGAAAG  223
               |||||||||||
Db    94  GlyProArg----------------ProAspProArgAlaLysProArgArgGly  105
Qy   224  CTCAGAGTGCCCCAAAGTGCCCACCCGCTGCCGCAGAAGACATGGAGATGGACAAAGCT  283
               :::         |:||||:::|||
Db   106  ValAlaArgIleGlnArgArgArgAlaArgArg-----------------GlnAla  120
Qy   284  CAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGTGTTCCAAAACGGTGGTCA  343
               :::||||||:::|||
Db   121  ArgArgArgArgProAlaGlyProAlaAlaThrArgLyArgArgAlaArg--ProHisArgGlnPro  139
Qy   344  GCAAGACTTACGAGAGAGAGGGACAGGGGACCCACCTCAGCCACCAGGTACGAGAGGGGATG  403
               :::|||        |:||||||
Db   140  GlyThrMetValArgLeuProGlyArgArgGlyGlyAlaProGlyAlaArgArg------  157
Qy   404  CTGGTGTGCTTGAACCTGGGCACCCAGAGAATGACATTGACAGAATCCTCCACAGCCA--  461
               :::         |||
Db   158  ------LeuProHisProAlaAlaAlaArgGlnProAlaProAlaAsp-ProProGlnProGl  175
```

```
Qy   462  --------------CGGCTCCCCAAC---GCGGAGGAGAAAATGTGCCAACC  496
                                |||    |||         :::
Db   462  yProArgGlnAlaLaGlyGlyLysArgGlnProGlyProGlyArgArgGlyArgGlyArgArgAlaArgI  195
Qy   497  TGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGAGCGAGAGGAGCCCACATGGAGGA  556
               ||||||
Db   195  eGlyGlnProAlaArgAlaGlyGlyAlaArgAlaAlaAspArgGlyLeuuAlaAlaAlaArgAlaArgAla--  214
Qy   557  GTGACAGCGGTAGACACCAGGAGCACAGGCGGCTGTATGGAGGGAGA------GGCTGAGGAGAGGC  610
               ||||||                                         |||||
Db   215  ---ArgArgAlaThrLeuGlyGlnArgGlyLeuArgArgHisAspGlyAlaGlyAlaGlyArgAr  233
Qy   611  CCGGACAGGATCGAGTGCAGGTGGCTCGTCGTGGTCAGGATCAAGCGCGCCCCTTGCCTCCCCAGG  670
               |||||    |||||||
Db   233  gArgGlyArgGlyTrpArg------------------------------------------  238
Qy   671  TAAACAGATTTACAGAGAAATCCAAAGCCAACACAGAAATATAGCCCAGTGGGCA  730
               :::|||         |::        |||
Db   239  -ArgGlnCysArgAlaArgProArgArg-GlyAlaAlaArgThrArgGlnProAlaAlaT  258
Qy   731  ACTTGAAAGGGAGGATGGCAGCAGTGGGCTGATGAACACATCCAGAGAGCTCAATC  790
               |||         :::        |::  :::|||
Db   258  -------VaIAlaProHisProAspProArgProAspArgArgAlaSerP  274
Qy   791  CTTTCAGTGAAGAGTTTGATTACGACGTGGCCATGTCACCCGCTACACAAAGGAGATG  850
               |||                  |||  |::|||
Db   274  ro---------------------AlaThrVaIAlaArgHisArgArgGlnT  284
Qy   851  AGGGCTATGGCCGCCCCCAAAGAAGGACCAAAACTGCTGAAAGGGCCAAGCGT  903
               ||||||:::|||
Db   284  hrGlyArgGlyAlaProGlyProGlyHisProGlyThrAlaGlyIlaGlyAlaArgArg  301

RESULT 10
US-09-252-991A-23771
;  Sequence 23771, Application US/09252991A
;  Patent No. 6551795
;  GENERAL INFORMATION:
;    APPLICANT:  Marc J. Rubenfield et al.
;    TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONA
;    TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;    FILE REFERENCE:  107196.136
;    CURRENT APPLICATION NUMBER:  US/09/252,991A
;    CURRENT FILING DATE:  1999-02-18
;    PRIOR APPLICATION NUMBER:  US 60/074,788
;    PRIOR FILING DATE:  1998-02-18
;    PRIOR APPLICATION NUMBER:  US 60/094,190
;    PRIOR FILING DATE:  1998-07-27
;    NUMBER OF SEQ ID NOS:  33142
;  SEQ ID NO 23771
;    LENGTH:  609
;    TYPE:  PRT
;    ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-23771

Alignment Scores:
Pred. No.:              0.00052           Length:          609
Score:                  126.50            Matches:         110
Percent Similarity:     32.66%            Conservative:    35
Best Local Similarity:  24.77%            Mismatches:      144
Query Match:            6.15%             Indels:          155
DB:                     4                 Gaps:            24

US-10-644-659A-1 (1-1146) x US-09-252-991A-23771 (1-609)

Qy     3  GGCTCCGGGCGAAAAGAAAGCGGGAGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATACG  62
               ||                                         |||
Db     3  ------------------------------------------ArgAlaThrArgGluArgThr  231
Qy    63  CAC------AGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCA--------  104
               |||                                      :::|||
Db   232  HisProAspProTyrProGlyHisArgTyrArgSerArgProSerAlaAspLeuAspArg  251
```

```
Qy    105  GTGGGGCAATGAGAAAGCATCAGGCAGGCCCAGAGGCCTACAGGCTGGCTGCCGGGAGG               164
                 ::                     |||||||      ||||||||
Db    252  ArgGlyArgGlyAlaLeuSerArgProGlyProGly--------LeuGlyArgGlyAla       268

Qy    165  GACCCAGACTCACCTCAAGCTCCTAAACCAATCACACCCCTACTTCACCAGAAAGC                224
           ||||||||
Db    269  AspProGly----------------------------------------------AlaGlyGly       274

Qy    225  TCAGAGTGCCCAAAGTCGCCACCCGCCTGCCAGAAGGACATGGACATGGACAAAGCTC               284
                                                     :::|||
Db    275  GlyArgCysProGlyArgAlaLeuProArgGlnArgPro----------GlyArgLeu       290

Qy    285  AGAGAAAGCCCCTGAGGTTTCCACATCAAAAGAAGAGAGTGTCCAAAACGGTGGTCAG               344
                                                            ||||
Db    291  ArgAlaLeuProProGlyGlyCysArgGlnAlaProGlyThrAlaAlaArgCysGlyAla       310

Qy    345  CAAGACTTACGAGAGAGGAGGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGC        404
           ||||||||  ||
Db    311  AspArgGlyArgIleArg------------------------------------------------   316

Qy    405  TGGTGTGTTGCTTCAACCTGGCAGCCAGAGAATGACATTGACGAGAATCCTCCACAGCCACGG       464
                                    |||  |||  ||||||||
Db    317  ------------------------------GlnArgProLeuGlnArgArg             323

Qy    465  CTCCCC-------------------------------------------                    470
           ||||||
Db    324  LeuProAlaArgArgGlyArgGlyGlyAlaLeuArgAlaGlyIleHisAlaGlyHisAlaGlyProAlaProGly   343

Qy    471  AACGCGGAGGAGAAATGTGCCAACTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGAT             530
                                                 :::
Db    344  ArgGlyGluGlyLyArgGlyAlaGlyArgProGly------ArgProAlaArgGlyAlaLeuProPro       361

Qy    531  GGAGCAGGAGGAGCCCACATGGAGGAGTGACACAGAGGACACGGCGGCTATGG                   590
           ||||||
Db    362  GlyAlaGlyAlaGlyGlyGlnProGluGlu---------------------------Gly   371

Qy    591  AGGAGAGGCTGAGGAGAGGCCCAGCAGGATGGAGTGCA-------------GGTGGC              635
           ||||||||            |||
Db    372  ArgArgGlyAlaGlySerAlaSerPheGlyTrpArgGlyAspSerLeuArgProGlyGly       391

Qy    636  TGTGGTCAGGATCAAGCGGCCCCCTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAA          695
               :::            |||||||                ||||||:::||||
Db    392  -LeuArgArg------ArgProAlaSerAlaHisArg-ThrGlyLeuArgArgArgAlaG            409

Qy    696  CTGCAAAGCCC--------AACAGAAATATAGCCCAGTGGCAACTTGAAAGGGAGAGATG           746
                   :::||
Db    409  lyGlyGlnProArgAspArgArgArgProLeuAla-----------------------           420

Qy    747  GCCAGCAGTGGGCTGATGAACACATACAATCCAGAAGCTCAATCCTTCAGTGAAGAGTT            806
           ||  |||||||
Db    421  ------------------AlaHisLeuProArgGlnHisLeuArgGlyAlaGlyArgLeuA          434

Qy    807  TGATTACCGAGCTCGGCCA-----------TGTCCACCCGCCTACACAAGAGGAGATGAGGG        854
              :::   ||||||                   |||
Db    434  rgArgSerGlyArgProGlnProAlaLeuCysAspAlaAlaAspGlyLeuArgThrArgA         454

Qy    855  CTATGGCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAGCGTGCTGAGGAGCA            914
           ||:||||||:::||:::|||||
Db    454  laGlyAlaAspProArgGlnProGln------------------ProGlyGlyLeuArgSer-       469

Qy    915  CATCTA---CAGGGAAATGAT-----------GGACATGTGCTTCATTATCTGCAC             956
           |||    |||                    |||
Db    470  HisAlaGlyAspGlyGluAsp----ProGlyLeuProAlaGlyGlyValAlaThrAlaProArg     489

Qy    957  AATGGCTCGCCACAGACGAGA----------TGGCAAGAT-------                      992
                                        |||||
Db    490  AsnGlyTrpProMetAlaSerGlyArgLeuTrpAsnGlyProGlnGlyProValProGly        509

Qy    993  TACTTTTGACGATCT-----CTTTGACAGATACGTTCGTATTTCAGATAAAGTAGTAGGG       1046
              ||||      ||||||              :::|||||:::||
Db    510  TyrArgTrpSerSerAspArgGlyLeuCysProArgSerGly-GlyLeuArgLysIleGl       529

Qy    1047 CATTCTCATGCGT---GCCAGGAAACATGGACTG--------                         1077
           |||||:::       |||||         :::
```

(right column)

```
Db    529  yProGluGlnArgValThrArgProHisGlyLeuAsnAlaIleGlySerGlyArgTyrAl    549
                     |||                           ||||||||||
Qy    1078 -GTAGACTTTGAAGGAGGAGATGCTA-----------------TGGCAAGGCCGGATGACCATGTTGT  1127
            :::||||||                           ||:||
Db    549  aValGluThrCysGlyLysIleValGlyArgTyrTrpArgSerProArgProArgGluHisAlaAlI   569

Qy    1128 GATT                                                           1131
           ::::
Db    569  eLeu                                                           570
```

RESULT 11
US-09-252-991A-27062
; Sequence 27062, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Marc J. Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 107196.136
;   CURRENT APPLICATION NUMBER: US/09/252,991A
;   CURRENT FILING DATE: 1999-02-18
;   PRIOR APPLICATION NUMBER: US 60/074,788
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: US 60/094,190
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
;   SEQ ID NO 27062
;   LENGTH: 741
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27062

Alignment Scores:
Pred. No.:   0.00081   Length: 741
Score: 125.00   Matches: 83
Percent Similarity: 36.14%   Conservative: 50
Best Local Similarity: 22.55%   Mismatches: 166
Query Match: 6.07%   Indels: 69
DB: Gaps: 13

US-10-644-659A-1 (1-1146) x US-09-252-991A-27062 (1-741)

```
Qy    13   GAAAAGGAAAGCGGGAGGGCCCAGCCAAGAGCGCCCTCCG------AAGTACGGCACA          66
                 :::::::||:::          ||||||||  :::             :::|||
Db    274  AspArgGlnGlnCysGlnGlyGlnValGlnSerThrValGluProAspArgArgArgThr      293

Qy    67   GCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCAATGAGAACCAGCATC      126
           ||  |||                ||                       ::: ||||
Db    294  AlaGluCysValGlyArgLeuArgArgGlnProGlyLyGlnHisGlnHisGlnAlaThr    313

Qy    127  AGGCAGGCCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGGACCACCTCACCTCAAGCT    186
                   |||         ||||
Db    314  GluGlnAlaHis---------------------------GlnLysValAlaGlnArg       323

Qy    187  CCTAAACCAATCACACCCCTACTTCACCAGAAAGCTCAGCTGCCCAAAGTCGCCCA        246
                 :::                     ::: :::
Db    324  GlnAlaProArgHisProAlaThrGlyHisProGlnArgArgGlnAlaGlyAlaGluVal    343

Qy    247  CCCCGGCCTGCCAGAAGGACATGGACATGGACAAAGCCCCCTGAGGTTTCT        306
           |||           ||         :::|||||:::||
Db    344  GlyThrGluHisGlyGlnGlyGlyAspGlyGlnInSerGlyLeuGlyInGly          363

Qy    307  CACATCAAAAAGAAAGAGAGTGTCAAAACGGTGGTCAGAACTTACGAGAGAGGA---    363
                  :::||||||
Db    364  -----GGGGAC----------------GTGAGCCACCTCAGCGCCACAGGTACGAG         396

Qy    364  GlnInGlyAspGlnArgIleAlaPheGluAlaValAspHisArgArgAsp-LysGlyTrpAr    403
                   ::::||||
Db    384  ArgHisGlnInTyrHisArgAspThrGlyMetAlaGlyProGlyGlnInArgGlyGlyGlu    383

Qy    397  AGGGATGCTGGTGTGCTTGAACCTGGCAGCCAGCAGGAATG----             436
           |||||:::                                :::
```

```
Query Match:   6.05%        Indels:  103
DB:            4            Gaps:    17

US-10-644-659A-1 (1-1146) x US-09-252-991A-25404 (1-256)

Qy   20 AAAGCGGGAGGGGCCCAGCCAAGAGCGCCCTCCGGAAGATACGGCCACCTGGTCA   79
         :::::|||||||  |||  :::
Db    1 ArgSerGlyArgThrProThrArgAlaSerGlyLyArgAsnArgArgAlaArgSer 20

Qy   80 TCA--------GCTTGGCCCGAGGTTGGCAGCAGTGGGGCGAATGAGAACAGCATCA  127
         |||
Db   21 GlyMetProCysArgArgTrpProAlaThrProAlaAlaArgHisTrpArg-----  37

Qy  128 GGCAGGCCCAGGAGCCTACAGGCTGCTGCCGCGGAGGGACCC----------AGGACT 175
         |||   ||| :::
Db   38 ---ArgLeuAlaArgAlaArgCysGlyCysGlyThrGlyValProAlaGlyTrpArgThr  56

Qy  176 CACCTCAAGTCTCAAACCAATCACACCCCTACTTCACCAGAAAGCTCAGA---GTG  232
         :::  :::||||  :::
Db   57 ArgArgTrpProValArgArgAlaAlaProProArgValArgggValPheArgAlaVal  76

Qy  233 CCCCAAAGTCGCCACCCGCCTGCCAGAGGACATGGAGGATGGCAGCAGTGG---     292
         ||||||||||||||
Db   77 ProProAlaArgHisProGlyAla---AsnGlyCysArgGlyAlaAlaAlaLeuArg  95

Qy  293 ACGAGAGAGGAGGGGACGTGAGCCACCTCAGCCACAGCACCGGTACGAGAGGGATGCTGGTGTGC  352
         :::
Db   96 ProValArg------GlyAlaProGlyArgTrpProCysArgIle 108

Qy  353 ACGAGAGAGGAGGGGACGTGAGCCACCTCAGCCACAGCACCGGTACGAGAGGGATGCTGGTGTGC  412
         ::: |||||||||
Db  109 ------AlaGlyArgGlyArgProGlyAlaProValAla 119

Qy  413 TTGAACCTGGGCAGCCAGGAGAATGACAGATTCCTCCAGCGCACGGCTCCCCAA  472
          ::|||:::|||
Db  120 GlyThrAlaGlyArgArgArgAlaThr-------AlaSerArgAlaArgProPro 136

Qy  473 CGCGGAGGAGAAAATGTGCCAACCTGGTCTTCGAGCTAACCAAGGGCTGGAGAGTGATGG  532
         |||:||:::|||
Db  137 ArgAlaAspArgGlyArgArgArgCysCysIleAla---GlyArgAlaGlyAla------ 153

Qy  533 AGCAGGAGGAGGCCCACATGGAGGCTGACAGCCTGACACAGAGGACAGGCGGCTATGGAG  592
         :::||||
Db  154 ---------GlyProProSerAlaAlaSerArg 161

Qy  593 GAGAGGCTGAGGAGAGGCCCAGCAGGATGGACTGCCAGGTGGCTGTGGTCAGGATCAAGC  652
         ||||||
Db  162 ArgValLeuArgGluSerProGlyArgArgGlyArgArgArgArgArgArgArgArgArg- 174

Qy  653 GCCCCTTGCCCTCCAGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGA  712
         ::|||||||:::
Db  175 ------ProProArg----------ValArgSerProProProArgAsnArg 187

Qy  713 AATATAGCCCAGTTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACACATAC  772
         ||| ||:::
Db  188 ArgCysAlaGlyTrpSerProThrArgAlaGlyAsp----------- 198

Qy  773 AATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCC  832
         |||||
Db  199 ArgProGlyAlaSer------------------------ArgProPro 206

Qy  833 GCCTACACAAAGGAGATGAGGGCTATGGCCGCCCCAAAAGAACCAAAACTGCTGAAA  892
         ||| :::
Db  207 ------------------ProArgAlaProProProAlaProArgSerArgArg 219

Qy  893 GGGCCAAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGAAATGATGCTTCATTATCT  952
         ||| :::
Db  220 Gly----------------------ThrSerSerArgTrpThr 226

Qy  953 GCACAATGG  961
         ::: |||
Db  227 AlaGlnTrp 229
```

```
Db  403 gGlyLeuAspLeuLeuGlnGlyMetAlaGluGluArgGlnArgGlnLysHisGlnAlaGl 423
Qy  437 ------------------------ACATTGACAGAATCCTCCACAGC  459
                                 ||||   ::::::
Db  423 uSerAspGlnArgAlaLeuAlaLysAlaLeuGlnAlaArgThrValAlaGlnAlaGluGlnGl 443

Qy  460 CACGGGTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTTCTGAGCTAACCAAGGGC  519
         ||| |||::: :::
Db  443 nGlnAlaGlyGlnGlnGly-GlnArgArgGlnArgGlyValGluAlaGluGlnLeuH 463

Qy  520 TGGAGAGTGATGGAGCAGAGGGAGAGGGACCCACATGGAGGCTGACAGCCTGACACAGAGGAC 579
         ||||||||    :::
Db  463 isHisGlnSerGlyAlaAspValGlyValGlyAlaGluHisHisCysGlnArgArgGlnAlaG 483

Qy  580 AGCGGCTATGGAGGAGGAGGCTGAGGAGAGAGGGCCCGAGCAGCAGGATGGACTGCCAGGTGCTGTG 639
         |||    :::
Db  483 inGlyAlaThrGly------GlyGluGluGlyArgCysHisGlnThrGlyValGlyGlyValAla 500

Qy  640 GTCAGGATCAAGCGCCCTTGCCCTCCAGGTAAACAGATTTACAGAGAAACTCAACTGC 699
         |||   ||:::||:::||
Db  500 laLeuGlnGlnAlaGlyAspAlaGluProGlyGluGln---GlyGlnProAlaValAlaG 519

Qy  700 AAAGCCCACAGAAATATAGCCCAGTTGAAAGGGAGATGGCAGCAGTGG---  756
         |||   ::||||||
Db  519 lnValAlaAlaGlu------ProValAlaGlyHisProGlyAla---GlnHisAla 536

Qy  757 GCTGATGAACACATACAATCC---CAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTAC 813
         |||   ::||||||||
Db  537 AlaGlyAspHisValArgAlaProGlnGlnValLysGlyHisGlyHisGlyGlnValAspProGln 556

Qy  814 GAGCCTGGCCATGTCCACCCGCCTACACAAAGGAGATGAGGGC----  855
         |||
Db  557 Glu------AspGlySerGlnHisHisGlyThrProGlySerGluTrpIleValLys 573

Qy  856 ------TATGGCCGCCCCAAAAGAACCAAAACTGCTGAAAGGGGCC  897
              :::||||||||:::|||
Db  574 GlyThrGlySerLeuThrHisGlyGlnProGlyArgGly------ThrValGluArgAla 591

Qy  898 AAGCGTGCTGAGGAGCACATCTACAGGGAAATGATGGAAATGATGCTTCATTATCTGCACA 957
         ::: |||||||||
Db  592 ArgAlaValGlnGlnValPheHisPheAspLeuAlaGlnLnAlaAlaAlaAlaGlyLyValCysAla 611

Qy  958 ATGGCTCGCCACAGACGA  975
         ||| :::
Db  612 HisAlaAspHisHisGln  617
```

```
Alignment Scores:
Pred. No.:                           Length:         256
Score:                   124.50      Matches:         78
Percent Similarity:       31.27%     Conservative:    23
Best Local Similarity:    24.15%     Mismatches:     119

Pred. No.:               0.000545
```

RESULT 13
US-09-252-991A-17580
; Sequence 17580, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17580
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17580

Alignment Scores:
Pred. No.:   0.000739
Score: 124.50
Percent Similarity: 35.85%
Best Local Similarity: 26.10%
Query Match: 6.05%
DB: 4

Length: 482
Matches: 107
Conservative: 40
Mismatches: 150
Indels: 113
Gaps: 22

US-10-644-659A-1 (1-1146) x US-09-252-991A-17580 (1-482)

Qy   14 AAAAGGAAAGCGGGAGGGAGGGCCCAAGAGAGGCGCCCTCCGGAAGATACGCACAGCCACCC 73
          :::|||| |||||||||| |||||| ::: |||
Db   86 ArgArgSerAlaGlnProAlaGlyProArgGlnAlaAlaAspArgArgGluHisProGly 105

Qy   74 TGGTCATCAGCTTGGCCCGAGTTGGCGCAGTGGCGAATGAGAACAGCATCAGGCAGG 133
          ||||| ||||||||| ||||| |||
Db 106 AsnAlaAlaValArgProGlyProGlyAlaGluGluGlyProGlyAlaGlyProArgSerHis 125

Qy 134 CCCAGGAGCCTACAGGCTGGCTGCCGG----------------------------------- 160
          ||||:||||||||||||||||||
Db 126 ---ArgAlaLeuArgAlaGlyAlaAlaGlyLeuArgArgGlnProArgProAlaGluGlyHisPro 144

Qy 161 --GAGGGACCCAGGAGCTCACCTCAAGTCCTAAACCAATCACACCCCTACTTCACACC 217
          ||| ::: ::: :::
Db 145 GlyValGlyArgArgAlaHisArgLeuProGlyAlaAspProAlaGlyProAspProGluSer 164

Qy 218 AGAAAGCTCAGAGTGCCCCAAAGTCGCCACCCCGCCTGCCAGAAGGACATGGAGATGGAC 277
          :::   :::
Db 165 ArgProGlyGluLeuProAlaValAlaHisAlaArgArgArgLeuIleGlyAsnHisAsp 184

Qy 278 AAAGCT----------------------CAGAGAAAGCCC-------------------- 295
          ::||| ||||||
Db 185 GluAlaProAlaAlaCysProValGlyAlaPhePheLeuProAlaArgLysProGlnAla 204

Qy 296 CTGAGGTTTCTC--- 307
          |||||
Db 205 LeuArgGlyLeuArgArgThrArgSerAlaArgArgAlaSerAsnArgGlnArgProThr 224

Qy 308 --ACATCAAAAAGAAAGAGAGGTGTCCAAAACCGTGGTCAGCAAGA------------- 349
          :::||| ||| ||||| ::::::|||
Db 225 SerThrSerAsnAlaGlnArgArgProIleSerArgArgProIleSerGlyGly 244

Qy 350 -----------------CTTACGAGAGGAGGGAGGGACCTGAGCCACCTCAGCCACA 388
          |||
Db 245 ProIleSerSerProProTyrProThrValGlu------ThrAlaAlaThrProIleThr 262

Qy 389 GGTACGAGA--GGGATGCTG-----GTGTGCTTGAACCTGGCAGCCAGGAGAATGACA 439
          ||||:::: |||||| :::||||||
Db 263 GlySerMetProGlyTrpLeuProAlaLeuIleThrIleGly----------Thr 278

Qy 440 TTGACAGAATCCTCCACAGCACCACGGCTCCCAACGCGGA-------------------GGA 481

; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 26942
; LENGTH: 437
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26942

Alignment Scores:
Pred. No.:   0.00142   Length: 437
Score: 121.50   Matches: 48
Percent Similarity: 37.85%   Conservative: 19
Best Local Similarity: 27.12%   Mismatches: 56
Query Match: 5.90%   Indels: 54
DB: 4   Gaps: 8

US-10-644-659A-1 (1-1146) x US-09-252-991A-26942 (1-437)

Qy   3 GGCTCCGGCGAAAAGGAAAGCGGGAGGGGCCCAGCCAAGAGCGCCTCCGGAAGAT--   59
            :::|||||| ||| ::: |||||||||||||||||||||||||
Db 250 GlyGlnGlyLyArgThrValAlaAspGlnProGlyAspHis 269

Qy   60 ------ACGACCAGCCACCTGGT-----------------------------------   77
         |||||
Db   60

Qy   78 --CATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGGAATGAGACAGCATCAGGCAGGC 134
           ||||| |||||||||||||||||
Db 270 GlyThrAspLeuAlaHisProHisProGlyLeuArgSerArgArgHisArgArg 289

Qy 135 CCAGGAGCCTACAGG-------CTGGTGCCCGGAGGGACCCAGGACTCACCTCAAGCTCC 188
         ||||| :::
Db 290 ProHisArgArgSerGlyArgThrValAlaAla------ProProArgArgThrArg 306

Qy 189 TAAACCAATCACACCCCCTACTTCACACGAAGAAGTCAGAGTGCCCCAAAGTCGCCACC 248
Db 307 ProGlyProHisArgAspProLeuAlaAlaGlyArgArgProGly-- 321

Qy 249 CCGCCTGCCAGGAGGACCATGGACAATGGACAAAGCTCAGAGAGAAAGCCCCTGAGGTTTCTCA 308
         |||||||| ::: |||||||||::: ||:||||
Db 322 ProProAlaProValArgAlaGlnGlyThrArgGlnArgAsnProTyrGlyArg 341

Qy 309 CATCAAAAGAAGAAGAGGTGTCAAAACGGTGGTTCAGCAAGACTTACGAGAGAGGGGA 368
               :::||| |||||||||||
Db 342 HisProGluGlnGlyProAlaAspArgGlyArgAspArgArg--- 360

Qy 369 CGGTGGCAGCCACCTCAG-------CCACAGGTACGAGGAGGGATGCTGGTGTGCTTGAACCTGG 422
         ||||||| ||:   :::|||
Db 361 ------ProAspGlnGlyIleProGlyProArgGlu-GlnProGl 376

Qy 423 GCAGCCAGAGAATGACATTGACAGAATCCTCACAGCCACGGCTCCCCA 471
         :::||||| ::::: ||| ::: ||:::|||
Db 376 yArgProGluLeuAsnValProSerMetGluAsnGluAsnGlyAspPro 392

Search completed: May 4, 2004, 08:56:46
Job time : 32 secs

US-10-644-659A-1 (1-1146) x US-09-252-991A-29215 (1-1113)

Qy   6 TCCGGGCGAAAAGGAAAGCGG----------GGAGGGCCAGCCAAGAGAGCGCCTCCG 53
            :::|||||| ||| :::
Db   81 AlaGlyArgThrGlyLeuArgHisGlnProHisGlyAsnAlaAlaGluArgProGly 100

Qy   54 GAAGATACGGCACAGCCAC---CCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGGC 110
            ::: |||||||:||||
Db 101 GlyAsnAlaAlaGlnHisArgProGlyHisArg------GlyArgThrAlaGluProGly 118

Qy 111 GAATGAGAA------CAGCATCAGGCAGGCCCCAGAGCCTACAGGCTGGCGTGCCCGGGAGG 164
               :::|||:::|||||ProHisArgAlaGluGlyAlaGlnGlyGly---ProGlyGl 137
Db 119 AsnHisArgAspProGluHisArgAlaGlyAlaGlyProGlyAla-GlnGlyGly---ProGlyGl 137

Qy 165 GACCCAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACACGAAGAAGC 224
             |||||||| |||
Db 137 yGluGlnProGlnValArgProArgGlnHisGluProArgAspProHisProAlaGl 157

Qy 225 TCAGAGTGCCCCAAAGTCGCCACCCCGCCTGCCAGGAGGACAATGGAGATGGACAAAGCTC 284
         ||||||| :::||| :::||| ArgAlaG 172
Db 157 n-ArgHisProArgPheHisLysProAlaAlaGluLys------ArgAlaG 172

Qy 285 AGAGAAAGCCCCTGAGGTTTCTCACAT---CAAAAGAAGAAGAGGTGTCCAAAACGGTGGT 341
         ||||| ||||:||:::
Db 172 ln-ProAlaProAlaGlyIleIleProHisAspHisProGluIledGlyGlyLysProAlaGly 191

Qy 342 ------CAGCAAGACTTACGA------GAGAGGGGGACGT 371
         |||||:::||| |||||||||
Db 192 AspHisGlnArgAspProArgPheLeuGluArgArgGlyArgGlnAlaGlySerGlyLys 211

Qy 372 ------GAGCCACCTCAGCCA-----------------------------------CAG 389
                :::|||:||
Db 211

Qy 390 GTACGAGAGGGGATGCTGGTGTGCTTGAACCTGGGCAGCCA---------------- 429
         :::|||
Db 212 ProProPheGlnSerProArgProAspProAspHisGlyArgProAspHisGlySerGlyArg 231

Qy 430 ------GAGAATGACATTGAC---------------------------------- 444
         |||||
Db 232 ProArgGluAla-ThrGlyThrGlyGlnProGlyLeuProGlyTyrProAspProIleAl 251

Qy 445 -------AGAATCCTCCACAGCCACGGCTCCCCAACGCGGGAGGAG 482
            :::|||
Db 251 aGlyArgProAlaAlaAlaAlaGluAlaAspProHisGlnProGlyArgGlnArgHisGlnVa 271

Qy 483 ------AAAATGT--------------------------------GCCAA 494
                |||||
Db 271 lHisProGlyArgHisArgArgArgAlaArgHisAlaArgArgLysArgArgProAr 291

Qy 495 CCTGGTGTCTGAGCTAACCAAGGGC------------------------TCG------ 522
         ||||||| |||
Db 291 gAlaAlaAlaAspGlnArgThrGlyTyrHisArgTy-ArgProLeuArgGlyGlyProAlaSe 311

Qy 523 --------------------AGAGTGATGGAGCAGGAGGAGGCCACATGGAGGAGGTGA 560
                             :::|||:::
Db 311 rLeuValGlnGlyLeuGlnProGlyArgGlnLeuThrValAlaAlaThrSerArgTrpHisAr 331

Qy 561 CAGCGTAGCACAGAGGACGACGCGGCTATGAGGAGGCTGAGGAGGCCCGAGCAGGA 620
         |||||:::::::::||:::|||
Db 331 gProGlyProGlyAspLeuGlnAlaProAspArgAlaAspArgArgAspArgArgAspArgArgAspArgArgAr 351

Qy 371 --------ArgArgGlnArgGlyAlaGlyLyArgLeuGlyLeuGlnArgGlyProThrArg- 386
Db 387 GlyAlaAlaArgThrAlaGlyThrArgAsp---AlaProLeuAlaAlaPro 401

Qy 621 TGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCTTGCCCTCCCA 668
         :::||||||||| |||||||

RESULT 15
US-09-252-991A-26942
; Sequence 26942, Application US/09252991A

OM nucleic - nucleic search, using sw model

```
          Copyright   GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: May 5, 2004, 00:51:08 ; Search time 525 Seconds
(without alignments)
9874.076 Million cell updates/sec

Title: US-10-644-659A-1
Perfect score: 1146
Sequence: 1 atggctccgggcgaaagga.........tgattacgctactcaagtga 1146

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2936184 seqs, 226173022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
c 1	1144.4	99.9	1322	16	US-10-104-047-1199	Sequence 1199, Ap
c 2	499.4	43.6	537	9	US-09-864-761-8525	Sequence 8525, Ap
c 3	363.4	31.7	365	9	US-09-864-761-25251	Sequence 25251, A
c 4	198.2	17.3	388	16	US-10-191-803-820	Sequence 820, App
5	42.8	3.7	594	13	US-10-142-426-10	Sequence 10, Appl
6	42.8	3.7	594	15	US-10-123-155-10	Sequence 10, Appl
7	42.8	3.7	594	15	US-10-146-731-10	Sequence 10, Appl
8	42.8	3.7	594	15	US-10-140-472-10	Sequence 10, Appl
9	42.8	3.7	594	15	US-10-141-761-10	Sequence 10, Appl
10	42.8	3.7	594	15	US-10-142-885-10	Sequence 10, Appl
11	42.8	3.7	594	15	US-10-158-790-10	Sequence 10, Appl
12	42.8	3.7	594	16	US-10-137-871-10	Sequence 10, Appl
13	42.8	3.7	594	16	US-10-140-923-10	Sequence 10, Appl
14	42.8	3.7	594	16	US-10-141-756-10	Sequence 10, Appl
15	42.8	3.7	594	16	US-10-141-759-10	Sequence 10, Appl
16	42.8	3.7	594	16	US-10-140-805-10	Sequence 10, Appl
17	42.6	3.7	594	16	US-10-140-864-10	Sequence 10, Appl
18	42.6	3.7	671	15	US-10-184-644-346	Sequence 346, App
19	42.6	3.7	671	15	US-10-184-634-346	Sequence 346, App
20	40.4	3.5	14800	9	US-09-954-456-1601	Sequence 1601, Appl
21	40.4	3.5	14800	9	US-10-269-909-61	Sequence 61, Appl
c 22	39.2	3.4	1046	13	US-10-425-114-23019	Sequence 23019, A
c 23	38.8	3.4	14835	15	US-10-240-965-113	Sequence 113, App
c 24	38.6	3.4	5751	17	US-10-641-643-1415	Sequence 1415, Ap
c 25	38.6	3.4	13210	10	US-09-814-353-22015	Sequence 22015, A
c 26	38.6	3.4	14255	9	US-09-964-824A-282	Sequence 282, App
c 27	38	3.3	732	13	US-10-424-599-29937	Sequence 29937, A
c 28	37.8	3.3	711	15	US-10-029-386-20329	Sequence 20329, A
c 29	37.8	3.3	1140	15	US-10-424-599-42070	Sequence 42070, A
30	37.4	3.3	486	9	US-09-938-842A-355	Sequence 355, App
31	37.4	3.3	486	11	US-09-938-842A-355	Sequence 355, App
c 32	37.4	3.3	492	9	US-09-770-961-618	Sequence 618, App
33	37.2	3.2	954	15	US-10-087-192-3934	Sequence 3934, Ap
c 34	37.2	3.2	235070	15	US-10-087-192-1990	Sequence 1990, Ap
35	37.2	3.2	9025608	15	US-10-156-761-1	Sequence 1, Appli
36	37	3.2	471	10	US-09-089-995-29236	Sequence 29236, A
37	37	3.2	1642	13	US-10-302-172-873	Sequence 873, App
38	36.4	3.2	999	15	US-10-184-644-434	Sequence 434, App
39	36.4	3.2	999	15	US-10-184-634-434	Sequence 434, App
c 40	36.4	3.2	1180	16	US-10-260-238-754	Sequence 754, App
41	36.2	3.2	660	13	US-10-142-426-480	Sequence 480, App
42	36.2	3.2	660	15	US-10-123-155-480	Sequence 480, App
43	36.2	3.2	660	15	US-10-146-731-480	Sequence 480, App
44	36.2	3.2	660	15	US-10-140-472-480	Sequence 480, App
45	36.2	3.2	660	15	US-10-141-761-480	Sequence 480, App

ALIGNMENTS

```
RESULT 1
US-10-104-047-1199
; Sequence 1199, Application US/1010047
; Publication No. US2003023 6392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1199
; LENGTH: 1322
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-1199

Query Match         99.9%;  Score 1144.4;  DB 16;  Length 1322;
Best Local Similarity 99.9%;  Pred. No. 0;
Matches 1145;  Conservative 0;  Mismatches 0;  Indels 1;  Gaps 0;

Qy     1 ATGGCTCCGGGCGAAAGGAAAAGGAAAAGCGGGGAGGGCCAAGAGCCAGCCAAGAGCGCCTCCGGAAGATA  60
           |||||||||||||||||||||||||||||| |||||||||||||||| |||||||||||||||||||||
Db    57 ATGGCTCCGGGCGAAAGGAAAAGGAAAAGTGGGGAGGGCCAAGAGCGCCAAGAGCGCCTCCGGAAGATA 116

Qy    61 CGGACAGCCACCCTGGTCATCAGCTTGGTCAGCAGGTTGGCAGCAGTGGGCGAATGAGAAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 CGGACAGCCACCCTGGTCATCAGCTTGGTCAGCAGGTTGGCAGCAGTGGGCGAATGAGAAC 176

Qy   121 AGGATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCGTCCGGGAGGAGGGACCCAGGACTCACCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 AGGATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCGTCCGGGAGGAGGGACCCAGGACTCACCT 236

Qy   181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCCAAAG 240
```

```
Db       |||||||||||||||||||||||||||||||||||||||||||
   237   CAAGCTCCTAAACCAATCACACCCCCACTTCACCACCAGAAAGCTCAGAGAAG  296

Qy       TCGCCACCCGCCTGCCAGAAGGACATGGAGAATGGACAAAGCTCAGAGAAAGCCCCTGAG  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241   TCGCCACCCGCCTGCCAGAAGGACATGGAGAATGGACAAAGCTCAGAGAAAGCCCCTGAG  356

Qy 301   GTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAGACTTACGAGAGA  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 357   GTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAGACTTACGAGAGA  416

Qy 361   GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCT  420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 417   GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCT  476

Qy 421   GGGCAGCAGCAGAGAATGACATTGACAGATACCTCCACAGCCACGGCTCCCCAACGCGGAGG  480
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 477   GGGCAGCAGCAGAGAATGACATTGACAGATACCTCCACAGCCACGGCTCCCCAACGCGGAGG  536

Qy 481   AGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG  540
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 537   AGAAAATGTGCCAACCCTGTGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG  596

Qy 541   GAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGACAGCGGCTATGGAGGAGAGAGCT  600
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 597   GAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGACAGCGGCTATGGAGGAGAGAGGCT  656

Qy 601   GAGGAGAGGCCCGAGCAGCAGGATGGAGTGCAGGTGCGTGTGGTCAGGATCAAGCCCCTTG  660
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 657   GAGGAGAGGCCCGAGCAGCAGGATGGAGTGCAGGTGCGTGTGGTCAGGATCAAGCCCCCTTG  716

Qy 661   CCCTCCCAGTGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  720
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 717   CCCTCCCAGTGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAAATATAGC  776

Qy 721   CCAGTGGCAACTTGAAAGGGAGAATGGCAGCAGGTGCTGATGAACATACAATCCCCAG  780
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 777   CCAGTGGCAACTTGAAAGGGAGAATGGCAGCAGGTGCTGATGAACATACAATCCCCAG  836

Qy 781   AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTTGGCCATGTCCACCCGCCTACAC  840
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 837   AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTTGGCCATGTCCACCCGCCTACAC  896

Qy 841   AAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGAACCAAAAACTGCTGAAAGGGCCAAG  900
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 897   AAAGGAGATGAGGGCTATGGCCGCCCCAAAGAAGAACCAAAAACTGCTGAAAGGGCCAAG  956

Qy 901   CGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG  960
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 957   CGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG  1016

Qy 961   GCTCGCCACACAGAGAATGGCAAGATCCAGTTACTTTGAGATCTCTTTGACAGATAC  1020
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1017  GCTCGCCACACAGAGAATGGCAAGATCCAGTTACTTTGAGATCTCTTTGACAGATAC  1076

Qy 1021  GTTCGTATTTCAGATAAAGTAGTGGGCATTCATCATGCGTGCCAGGAAACATGGACTGGTA  1080
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1077  GTTCGTATTTCAGATAAAGTAGTGGGCATTCATCATGCGTGCCAGGAAACATGGACTGGTA  1136

Qy 1081  GACTTTGAAGGAGGAGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTGATTACGCTACTC  1140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1137  GACTTTGAAGGAGGAGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTGATTACGCTACTC  1196

Qy 1141  AAGTGA  1146

Db 1197  AAGTGA  1202
```

RESULT 2
US-09-864-761-8525/c
; Sequence 8525, Application US/09864761
; Patent No. US2002004876341
; GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FC
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 8525
  LENGTH: 537
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
  OTHER INFORMATION: MAP TO AC023344.2
  OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
  OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
  OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
  OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
  OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
  OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
US-09-864-761-8525

Query Match   43.6%;   Score 499.4; DB 9; Length 537;
Best Local Similarity 99.8%; Pred. No. 6.1e-152;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 169 CAGGACTCACCTCAGCTCCTAAACCAATCACACCCCCACTTCACCACCAGAAAGCTCAG 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 537 CAGGACTCACCTCAGCTCCTAAACCAATCACACCCCCACTTCACCACCAGAAAGCTCAG 478

Qy 229 AGTGCCCAAAGTGCCACCCGCTGCCACCCGCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAG 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 477 AGTGCCCAAAGTCGGCACCCGCTGCCACCCGCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAG 418

```
Qy  289  AAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAG  348
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  417  AAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAG  358

Qy  349  ACTTACGAGAGAGGAGGGGACGTGAGCCACCTCAGCCCACCAGGTACGAGAGGGATGCTGGT  408
         |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  357  ACTTATGAGAGAGGAGGGGACGTGAGCCACCTCAGCCCACCAGGTACGAGAGGGATGCTGGT  298

Qy  409  GTGCTTGAACCTGGCAGCCAGCGAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC  468
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  297  GTGCTTGAACCTGGCAGCCAGCGAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC  238

Qy  469  CCAACGCGGAGGAGAAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGAGAGTG  528
         |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  237  CCAACGCGGAGGAGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGAGAGTG  178

Qy  529  ATGGAGCAGGAGGAGGAGCGCCACATGAGGAGGGCTAGACACAGAGGACACAGCGGGCTAT  588
         ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  177  ATGGAGCAGCAGGAGGAGCGCCACATGAGGAGGGCTAGACACAGAGGACACAGCGGGCTAT  118

Qy  589  GGAGGAGAGGCTGAGGAGGAGCCCGAGCAGGATGGAGTGCAGGTGCGTGTGGTCAGGATC  648
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117  GGAGGAGAGGCTGAGGAGGAGCCCGAGCAGGATGGAGTGCAGGTGCGTGTGGTCAGGATC  58

Qy  649  AAGCGCCCCTTGCCCTCCCAG  669
         |||||||||||||||||||||
Db  57   AAGCGCCCCTTGCCCTCCCAG  37
```

RESULT 3
US-09-864-761-25251/c
; Sequence 25251, Application US/09864761
; Patent No. US2002004876 3A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25251
; LENGTH: 365
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023344.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.8
; OTHER INFORMATION: NT HIT: AJ250189.1, EVALUE 2.30e-01
; OTHER INFORMATION: SWISSPROT HIT: P09933, EVALUE 1.00e+01
US-09-864-761-25251

Query Match   31.7%;   Score 363.4; DB 9;   Length 365;
Best Local Similarity   99.7%;   Pred. No. 1.2e-107;
Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy  260  AAGGACATGGAGATGGACAAAGTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA  319
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  365  AAGGACATGGAGATGGACAAAGTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGA  306

Qy  320  AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACGTGAGCCACC  379
         |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  305  AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGGAGGGGACGTGAGCCACC  246

Qy  380  TCAGCCCACCAGGTACGAGAGGGATGCTGGTGCTTGAACCTGGCAGCCAGCGAGAATGACA  439
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  245  TCAGCCCACCAGGTACGAGAGGGATGCTGGTGCTTGAACCTGGCAGCCAGCGAGAATGACA  186

Qy  440  TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAAATGTGCCAACCTGG  499
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  185  TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAAATGTGCCAACCTGG  126

Qy  500  TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGTGATGGAGCCCATGAGGAGGAGTG  559
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  125  TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGTGATGGAGCCCATGAGGAGGAGTG  66

Qy  560  ACAGGGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGGAGGAGCCCGAGCAGG  619
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65   ACAGGGTAGACACAGAGGACAGCGGCTATGGAGGAGGCTGAGGAGGAGCCCGAGCAGG  6

Qy  620  ATGGA  624
         |||||
Db  5    ATGGA  1
```

RESULT 4
US-10-191-803-820/c
; Sequence 820, Application US/10191803
; Publication No. US20040014040A1
; GENERAL INFORMATION:
; APPLICANT: MENDRICK, Donna
; APPLICANT: PORTER, Mark
; APPLICANT: JOHNSON, Kory
; APPLICANT: HIGGS, Brandon
; APPLICANT: CASTLE, Arthur
; APPLICANT: ELASHOFF, Michael
; TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
; FILE REFERENCE: 4921-509US
; CURRENT APPLICATION NUMBER: US/10/191,803

; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/303,819
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,623
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/369,351
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: US 60/377,611
; PRIOR FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 1140
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 820
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040014040A1 AI172339
US-10-191-803-820

Query Match   17.3%; Score 198.2; DB 16; Length 388;
Best Local Similarity 88.5%; Pred. No. 9.4e-54;
Matches 215; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

Qy   904 GCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATGGCT 963
          ||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||
Db   347 GCCGAGGAGCACATCTATCGGGAAATTATGATGGAATTGTGCTTTGTTATCCGCACAATGGCT 288

Qy   964 CGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGACAGATACGTT 1023
          ||||||||||||||||||||||||||||||||||||| |||||||||||| |||||||||
Db   287 CGCCATAGACGAGATGGCAAGATCCAGGTTACTTTCGGAGAACTTCTTTGATAGATATGTT 228

Qy   1024 CGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGAGAAACATGGACTGGTAGAC 1083
          ||||||||||||||||||||||||||||| | |||||||||||||| |||||||||||
Db   227 CGAATTTCAGATAAAGTGGTGGGCATCCTCATGCGTGCCAGAGAACACGGACTGGTGCAC 168

Qy   1084 TTTGAAGGAGGAGATGCTATGGCAAGGCCCGAGATGACCATGTTGTGATTACGCTACTCAAG 1143
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   167 TTTGAAGGGGGAGATGCTATGGCAAGGCAAAGATGACCATGTTGATTACTCTCCCTTGAG 108

Qy   1144 TGA 1146
          |||
Db   107 TAA 105

RESULT 5
US-10-142-426-10
; Sequence 10, Application US/10142426
; Publication No. US2004004833A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C224
; CURRENT APPLICATION NUMBER: US/10/142,426
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-426-10

Query Match   3.7%; Score 42.8; DB 13; Length 594;
Best Local Similarity 10.0%; Pred. No. 0.0057;
Matches 37; Conservative 152; Mismatches 177; Indels 3; Gaps 1;

Qy   391 TACGAGAGGGATGCTGGTGTGTCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC 450
          ||:|.|:| :: .:: .:| :::|::|| :: :::::: :: .:|: .|:::::
Db   226 TTYGMADGDFVRMLRKDKEEEAIKHAKALEEKAMYSGRRSRRQRREFREKRLRGRKIS 285

Qy   451 CTCCACAGCCACGGCTCCCCAACGCCGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA 510
          :::|:|:| :::|::: |:::: :|: :::::|:: :| :::: ::|: :|:::|—|
Db   286 PPSYARRDSPTYDPYKRSPSESSSESRSRSRSRPTPGREEKITFITSFGGSDEEAAAAAA 345

Qy   511 ACCAAGGGCTGAGAGAGTGATGAGCAGGAGGAGGGCCCACATGGAGGAGTGACAGCGTAGAC 570
          ::| ::: ::|: :::: ::::|::|:: |:||| :: .:: ::
Db   346 AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSSASRTSSSSRSSRSSSRSRRG 405

Qy   571 ACAGAGGACAGCCGGCTATGGAGGAGGAGGGCTGAGGAGGAGGGCCGAGCAGGATGGAGTGC—- 628
          :|:|:: |::|:::|:||:||:||:||::: :::|::| :::::|: ::|:|:|: ||
Db   406 GGYRYSGRHARRSRWSRSRRSRRYSRSRSRGRHRHSGGGSRDGHRYRSRSPARRGGYGP 465

Qy   629 -AGGTGGCTGTGGTCAGGATCAAGCGCCCTTGCCCTCCCAGGTAAACAGATTTACAGAG 687
          :::: :: :: ::|:::| : ::::||:|::|: |:|::: ||:|:||||:|:|
Db   466 RRRSRSRSHSGDRYRRGGRGLRHHSSSRSRSRSWSLSPSRSRSLTRSRSHSPSPSQSRSRS 525

Qy   688 AAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGGATGG 747
          ::: | |:| |: |: ::| :::||| :|:|: :||:|
Db   526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy   748 CAGCAGTGG 756
          |::|:||||
Db   586 GAEQAVQGG 594

RESULT 6
US-10-123-155-10
; Sequence 10, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-123-155-10

```
Query Match         3.7%;   Score 42.8;  DB 15;  Length 594;
Best Local Similarity  10.0%;  Pred. No. 0.0057;
Matches  37;  Conservative 152;  Mismatches 177;  Indels  3;  Gaps  1;

Qy    391  TACGAGAGGGATGCTGGTGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC  450
               |:|:|:|:|:| :::  ::::: : :  :::::: :  |::::: :: ::|::|
Db    226  TTYGMADGDFVRMLRKDKEEABAIKHAKALEEEKAMYSGRRSRRQRREFEKRLRGRKIS    285

Qy    451  CTCCACAGCCACGGCTCCCCAACGCGGAGAGAAAATGCCAACCTGGTGTCTGAGCTA    510
               ::: ::::::: :: ::::: : :::: :  ::::  :::::: :
Db    286  PPSYARRDSPTYDPYKRSPSESSSESRSRSRSPTPGREEKITFITSFGGSDEAAAAAA   345

Qy    511  ACCAAGGGCTGGAGAGTGATGGAGCAGGAGGCCCACATGGAGGAGTGACAGCGTAGAC    570
               ::: ::::::::|: ::  :::: ::::::::::: :|: |
Db    346  AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSASRTSSSRSSSRSSSRRG    405

Qy    571  ACAGAGACACGGCGGCTATGGAGCAGGAGGCTGAGGAGGCCCGAGCAGGATGGAGTGC--   628
               :||:|||::| |: ::::: :::: | :|:: :|::::|  |::|::: |:
Db    406  GGYYRSGRHARSRSRSWSRSRSRRYSRSRSRGRRHSGGGSRGDGHRYSRSPARRGGYGP   465

Qy    629  -AGGTGGCTGTGTCAGGATCAAGCGCCCCTTGCCTCCAGGTAAACAGATTTACAGAG    687
               :::  ::::|::  ::: :::| ::::::: ::::| :::::: :::::::
Db    466  RRRSRSRSHSGDRYRRGGRGLRHHSSSRSRSSWSLSPSRSRSLTRSRSHSPSPSQSRSRS   525

Qy    688  AAACTCAACTGCAAAGCCCAACAGAAAATAGCCCAGTGGGCAACTTGAAAGGGAGATGG  747
               ::: ::::|::::: :::: :||:::|::::::|: |:::::|
Db    526  RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE  585

Qy    748  CAGCAGTGG  756
               | |:|
Db    586  GAEQAVQGG  594
```

RESULT 7
US-10-146-731-10
; Sequence 10, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang,Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3330R1C323
;   CURRENT APPLICATION NUMBER: US/10/146,731
;   CURRENT FILING DATE: 2002-05-15
;   PRIOR Application removed - See File Wrapper or Palm
;   NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-146-731-10

```
Query Match         3.7%;   Score 42.8;  DB 15;  Length 594;
Best Local Similarity  10.0%;  Pred. No. 0.0057;
Matches  37;  Conservative 152;  Mismatches 177;  Indels  3;  Gaps  1;
```

```
Qy    391  TACGAGAGGGATGCTGGTGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC  450
               |:|:|:|:|:| :::  ::::: : :  :::::: :  |::::: :: ::|::|
Db    226  TTYGMADGDFVRMLRKDKEEABAIKHAKALEEEKAMYSGRRSRRQRREFEKRLRGRKIS    285

Qy    451  CTCCACAGCCACGGCTCCCCAACGCGGAGAGAAAATGCCAACCTGGTGTCTGAGCTA    510
               ::: ::::::: :: ::::: : :::: :  ::::  :::::: :
Db    286  PPSYARRDSPTYDPYKRSPSESSSESRSRSRSPTPGREEKITFITSFGGSDEAAAAAA   345

Qy    511  ACCAAGGGCTGGAGAGTGATGGAGCAGGAGGCCCACATGGAGGAGTGACAGCGTAGAC    570
               ::: ::::::::|: ::  :::: ::::::::::: :|: |
Db    346  AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSASRTSSSRSSSRSSSRRG    405

Qy    571  ACAGAGACACGGCGGCTATGGAGCAGGAGGCTGAGGAGGCCCGAGCAGGATGGAGTGC--   628
               :||:|||::| |: ::::: :::: | :|:: :|::::|  |::|::: |:
Db    406  GGYYRSGRHARSRSRSWSRSRSRRYSRSRSRGRRHSGGGSRGDGHRYSRSPARRGGYGP   465

Qy    629  -AGGTGGCTGTGTCAGGATCAAGCGCCCCTTGCCTCCAGGTAAACAGATTTACAGAG    687
               :::  ::::|::  ::: :::| ::::::: ::::| :::::: :::::::
Db    466  RRRSRSRSHSGDRYRRGGRGLRHHSSSRSRSSWSLSPSRSRSLTRSRSHSPSPSQSRSRS   525

Qy    688  AAACTCAACTGCAAAGCCCAACAGAAAATAGCCCAGTGGGCAACTTGAAAGGGAGATGG  747
               ::: ::::|::::: :::: :||:::|::::::|: |:::::|
Db    526  RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTPAAGKETGAAKVTQADASGEAETEDAE  585

Qy    748  CAGCAGTGG  756
               | |:|
Db    586  GAEQAVQGG  594
```

RESULT 8
US-10-140-472-10
; Sequence 10, Application US/10140472
; Publication No. US2003013888A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang, Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3330R1C168
;   CURRENT APPLICATION NUMBER: US/10/140,472
;   CURRENT FILING DATE: 2002-05-06
;   Prior Application removed - See File Wrapper or Palm
;   NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-472-10

```
Query Match         3.7%;   Score 42.8;  DB 15;  Length 594;
Best Local Similarity  10.0%;  Pred. No. 0.0057;
Matches  37;  Conservative 152;  Mismatches 177;  Indels  3;  Gaps  1;

Qy    391  TACGAGAGGGATGCTGGTGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC  450
               |:|:|:|:|:| :::  ::::: : :  :::::: :  |::::: :: ::|::|
Db    226  TTYGMADGDFVRMLRKDKEEAIKHAKALEEEKAMYSGRRSRRQRREFEKRLRGRKIS    285

Qy    451  CTCCACAGCCACGGCTCCCCAACGCGGAGAGAAAATGCCAACCTGGTGTCTGAGCTA    510
```

Left column (RESULT 9)

```
Db  286  ::|::  :: ::|: :  :: :  :  :   : : : ::::|      |::|:|       |
         PPSYARRDSPTYDPYKRSPSESSESRSRSRSPTPGREEKITFITSFGGSDEEAAAAAAA  345

Qy  511  ACCAAGGGCTGGAGAGTGATGAGGAGAGGCTCCACATGGCAGCCAGAGAATGACCAGCGTAGAC  570
         : :  ||  :|:| ::: ::|: : :||: :   ::  :  :  : |:  : ::::

Db  346  AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSSASRTSSRSSSRSSSSRSRRG  405

Qy  571  ACAGAGGACAGCGGCTATGAGGAGAGGCTCGAGGAGAGCCCGAGCAGGATGGAGTGC--  628
         :|:::|:|  ::  |::: : ::  : ::::  : : ::::|   ::: :|::|

Db  406  GGYYRSGRHARSRSWSRSRSRRYSRSRSRSRGRRHSGGGSRDGHRYSRSPARRGGYGP  465

Qy  629  -AGGTGGCTGTGGTCAGGATCAAGGCGCCCCTTGCCTCCCAGGTAAACAGATTTACAGAG  687
          :|:|:::  :: :::  |||:  :  ::  : :  |:|:| :   ::  :: ::

Db  466  RRSRSRSHSGDRYRRGRGRLRHHSSSRSRSSWSLSPSRSRSLITRSRSHSPSPSQSRSRS  525

Qy  688  AAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGG  747
         ::  : ::  ::  |||: :: ::|      :||| ::| ::  |::|   ::| :|

Db  526  RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE  585

Qy  748  CAGCAGTGG  756
         |::|  ||

Db  586  GAEQAVQGG  594
```

RESULT 9
US-10-141-761-10
; Sequence 10, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-761-10

```
Query Match             3.7%;  Score 42.8;  DB 15;  Length 594;
Best Local Similarity  10.0%;  Pred. No. 0.0057;
Matches   37;  Conservative  152;  Mismatches  177;  Indels    3;  Gaps    1;

Qy  391  TACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC  450
         ::|  :|| :  ::  :||:|     ::  ::|:

Db  226  TTYGMADGDFVRMLRKDKEEAEAIKHAKALEEKAMYSGRRSRRQRREFREKRLRGRKIS  285

Qy  451  CTCCACAGCCACGGCTCCCCAACGGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA  510
         : ::  |: : :: : :: ::  :::|::|:  :: |:|:|

Db  286  PPSYARRDSPTYDPYKRSPSESSESRSRSRSPTPGREEKITFITSFGGSDEEAAAAAAA  345

Qy  511  ACCAAGGGCTGGAGAGTGATGAGGAGAGGCTCCACATGGCAGCCAGAGAGTGAC  570
         : :  ||  :|:| ::: ::|: : :||: :   ::
```

Right column (RESULT 10)

```
Db  346  AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSASRTSSRSSSRSSSSRSRRG  405

Qy  571  ACAGAGGACAGCGGCTATGAGGAGAGGCTCGAGGAGAGCCCGAGCAGGATGGAGTGC--  628
         :|:::|:|  ::  |::: : ::  : ::::  : : ::::|   ::: :|::|

Db  406  GGYYRSGRHARSRSWSRSRSRRYSRSRSRSRGRRHSGGGSRDGHRYSRSPARRGGYGP  465

Qy  629  -AGGTGGCTGTGGTCAGGATCAAGGCGCCCCTTGCCTCCCAGGTAAACAGATTACAGAG  687
          :|:|:::  :: :::  |||:  :  ::  :  |:|:| :   ::  :: ::

Db  466  RRSRSRSHSGDRYRRGRGRLRHHSSSRSRSSWSLSPSRSRSLITRSRSHSPSPSQSRSRS  525

Qy  688  AAACTCAACTGCAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGG  747
         ::  : ::  ::  |||: :: ::|      :||| ::| ::  |::|   ::| :|

Db  526  RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE  585

Qy  748  CAGCAGTGG  756
         |::|  ||

Db  586  GAEQAVQGG  594
```

RESULT 10
US-10-142-885-10
; Sequence 10, Application US/10142885
; Publication NO. US2003015760A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C248
; CURRENT APPLICATION NUMBER: US/10/142,885
; CURRENT FILING DATE: 2002-05-10
; Prior Apploication removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-142-885-10

```
Query Match             3.7%;  Score 42.8;  DB 15;  Length 594;
Best Local Similarity  10.0%;  Pred. No. 0.0057;
Matches   37;  Conservative  152;  Mismatches  177;  Indels    3;  Gaps    1;

Qy  391  TACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC  450
         ::|  :|| :  ::  :||:|     ::  ::|:

Db  226  TTYGMADGDFVRMLRKDKEEAEAIKHAKALEEKAMYSGRRSRRQRREFREKRLRGRKIS  285

Qy  451  CTCCACAGCCACGGCTCCCCAACGGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA  510
         : ::  |: : :: : :: ::  :::|::|:  :: |:|:|

Db  286  PPSYARRDSPTYDPYKRSPSESSESRSRSRSPTPGREEKITFITSFGGSDEEAAAAAAA  345

Qy  511  ACCAAGGGCTGGAGAGTGATGAGGAGAGGCTCCACATGGCAGCCAGAGAGTGACAGCGTAGAC  570
         : :  ||  :|:| ::: ::|: : :||: :   ::  :  :  : |:  : ::::

Db  346  AAASGVTTGKPPAPPQPGGPAPGRNASARRRSSSSSSSSSASRTSSRSSSRSSSSRSRRG  405

Qy  571  ACAGAGGACAGCGGCTATGAGGAGAGGCTCGAGGAGAGCCCGAGCAGGATGGAGTGC--  628
         :|:::|:|  ::  |::: : ::  : ::::  : : ::::|   ::: :|::|

Db  406  GGYYRSGRHARSRSWSRSRSRRYSRSRSRSRGRRHSGGGSRDGHRYSRSPARRGGYGP  465
```

```
Qy  629 -AGGTGGCTGTGTGTCAGGATCAAGCCGCCCCTTGCCCTCCCAGTAAACAGATTTACAGAG 687
         :::: :: :::::|| : ::: :::  ::::|| |::::
Db  466 RRRSRSRSHSGDRYRRGRGLRHHSSSRSRSWSLSPSRSRSLTRSRSHSPSPSQSRSRS 525

Qy  688 AAACTCAACTGCAAAGCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGG 747
         : .:: :: |||:|:|:  :::|:::|| |||: ||||||
Db  526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy  748 CAGCAGTGG 756
         | |:||
Db  586 GAEQAVQGG 594
```

RESULT 11
US-10-158-790-10
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C448
; CURRENT APPLICATION NUMBER: US/10/158,790
; CURRENT FILING DATE: 2002-05-30
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-158-790-10

Query Match   3.7%; Score 42.8; DB 15; Length 594;
Best Local Similarity 10.0%; Pred. No. 0.0057;
Matches   37; Conservative 152; Mismatches 177; Indels 3; Gaps 1;

```
Qy  391 TACGAGAGGGATGCGTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC 450
         |:|: :|:: :::
Db  226 TTYGMADGDFVRMLRKDKEEABAIKHAKALEEEKAMYSGRRSRRQRREFREKRLRGRKIS 285

Qy  451 CTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA 510
         ::||: ||  : ::::::|| ::||:  ::::::||:|
Db  286 PPSYARRDSPTYDPYKRSPSESSSESRSRSRSPTPGREEKITFITSFGGSDEEAAAAAAA 345

Qy  511 ACCAAGGGCTGGAGAGTGATGATGGAGCCCCATGGAGAGTGACAGCGTAGAC 570
         :|:| :: :| :: :::|| :|| : ::::: |:
Db  346 AAASGVTTGKPAPPQPGGPAPGRNASARRRSSSSSSSASRTSSSSRSSSSRSRRG 405

Qy  571 ACAGAGGACAGCGGCTATGGAGGAGAGGCTGAGGAGGCCCGAGCAGGATGGAGTGC-- 628
         |:  :::  ::: :::::: : ::::  ::: : :::||:
Db  406 GGYYRSGRHARSRSWSRSRSRRYSRSRGRRHSGGSRDGHRYSRSPARRGGYGP 465

Qy  629 -AGGTGGCTGTGTGTCAGGATCAAGCCGCCCCTTGCCCTCCCAGTAAACAGATTTACAGAG 687
         :::: :: :::::|| : ::: :::  ::::|| |::::
Db  466 RRRSRSRSHSGDRYRRGRGLRHHSSSRSRSWSLSPSRSRSLTRSRSHSPSPSQSRSRS 525
```

```
Qy  688 AAACTCAACTGCAAAGCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGG 747
         : .:: :: |||:|:|:  :::|:::|| |||: ||||||
Db  526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy  748 CAGCAGTGG 756
         | |:||
Db  586 GAEQAVQGG 594
```

RESULT 12
US-10-137-871-10
; Sequence 10, Application US/10137871
; Publication No. US20030207350A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-871-10

Query Match   3.7%; Score 42.8; DB 16; Length 594;
Best Local Similarity 10.0%; Pred. No. 0.0057;
Matches   37; Conservative 152; Mismatches 177; Indels 3; Gaps 1;

```
Qy  391 TACGAGAGGGATGCGTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC 450
         |:|: :|:: :::
Db  226 TTYGMADGDFVRMLRKDKEEABAIKHAKALEEEKAMYSGRRSRRQRREFREKRLRGRKIS 285

Qy  451 CTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA 510
         ::||: ||  : ::::::|| ::||:  ::::::||:|
Db  286 PPSYARRDSPTYDPYKRSPSESSSESRSRSRSPTPGREEKITFITSFGGSDEEAAAAAAA 345

Qy  511 ACCAAGGGCTGGAGAGTGATGATGGAGCCCCATGGAGAGTGACAGCGTAGAC 570
         :|:| :: :| :: :::|| :|| : ::::: |:
Db  346 AAASGVTTGKPAPPQPGGPAPGRNASARRRSSSSSSSASRTSSSSRSSSSRSRRG 405

Qy  571 ACAGAGGACAGCGGCTATGGAGGAGAGGCTGAGGAGGCCCGAGCAGGATGGAGTGC-- 628
         |:  :::  ::: :::::: : ::::  ::: : :::||:
Db  406 GGYYRSGRHARSRSWSRSRSRRYSRSRGRRHSGGSRDGHRYSRSPARRGGYGP 465

Qy  629 -AGGTGGCTGTGTGTCAGGATCAAGCCGCCCCTTGCCCTCCCAGTAAACAGATTTACAGAG 687
         :::: :: :::::|| : ::: :::  ::::|| |::::
Db  466 RRRSRSRSHSGDRYRRGRGLRHHSSSRSRSWSLSPSRSRSLTRSRSHSPSPSQSRSRS 525

Qy  688 AAACTCAACTGCAAAGCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGG 747
         : .:: :: |||:|:|:  :::|:::|| |||: ||||||
Db  526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy  748 CAGCAGTGG 756
```

```
Db        586 GAEQAVQGG 594
                  |:||

RESULT 13
US-10-140-923-10
; Sequence 10, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-923-10

  Query Match             3.7%;  Score 42.8;  DB 16;  Length 594;
  Best Local Similarity  10.0%;  Pred. No. 0.0057;
  Matches    37; Conservative  152; Mismatches  177;  Indels    3;  Gaps    1;

Qy    391 TACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC 450
          ||:|:|| :|:: :|:|  |||: :||||||||||:::::| :: :|: || :|:
Db    226 TTYGMADGDFVRMLRKDKEEAEAIKHAKALEEEKAMYSGRRSRQRREFREKRLRGRKIS 285

Qy    451 CTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA 510
          ||| ::: | |||| ::: :|| |::| ||| :| ||||||||:||
Db    286 PPSYARRDSPTYDPYKRSPSESSSESRSRSPTPGREEKITFITSFGGSDEEAAAAAAA 345

Qy    511 ACCAAGGGCTGGAGAGTGATGGAGCAGGAGGCCCACATGGAGGAGTGACAGCGTAGAC 570
          || :: :|||||||||||||: :|||:||||||||:: ::||::| :||:||||
Db    346 AAASGVTTGKPAPPQPGGPAPGRNASARRRSSSSSSASRTSSSRSSRSSSSRSRRG 405

Qy    571 ACAGAGGACAGCCGCTATGGAGCAGGAGGCTGAGGAGAGGCCCAGACGATGGAGTGC-- 628
          || ::|: ::|: | :::|:| ||| | ::::||||||: :|:| |||: :|| ::
Db    406 GGYYRSGRHARSRSRSWSRSRSRSRRYSRSRSRGRRHSGGGSRDGHRYSRSPARRGGYGP 465

Qy    629 -AGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCTCCAGGTAAACAGATTTACAGAG 687
          ::: ::|::|: |::|| ::: :| :|| ||::::|:|| ::: :|:| |::|:| :
Db    466 RRRSRSRSHSGDRYRRGGRGLRHHSSSRSSRSSWSLSPSRSRSLITRSRSHSPSQSRSRS 525

Qy    688 AAACTCAACTGCAAAGCCAACAGAAATATAGCCCAGAAAGGGAGATGG 747
          :::: :|::: |: :: :| :::|||||:||| :::|| ||:|::|:|
Db    526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy    748 CAGCAGTGG 756
          |:|:|||||
Db    586 GAEQAVQGG 594

RESULT 14
```

```
US-10-141-756-10
; Sequence 10, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
;   APPLICANT: Beresini,Maureen
;   APPLICANT: DeForge,Laura
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Filvaroff,Ellen
;   APPLICANT: Gao,Wei-Qiang
;   APPLICANT: Gerritsen,Mary E.
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Sherwood,Steven
;   APPLICANT: Smith,Victoria
;   APPLICANT: Stewart,Timothy A.
;   APPLICANT: Tumas,Daniel
;   APPLICANT: Watanabe,Colin K
;   APPLICANT: Wood,William
;   APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10/141,756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-756-10

  Query Match             3.7%;  Score 42.8;  DB 16;  Length 594;
  Best Local Similarity  10.0%;  Pred. No. 0.0057;
  Matches    37; Conservative  152; Mismatches  177;  Indels    3;  Gaps    1;

Qy    391 TACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATC 450
          ||:|:|| :|:: :|:|  |||: :||||||||||:::::| :: :|: || :|:
Db    226 TTYGMADGDFVRMLRKDKEEAEAIKHAKALEEEKAMYSGRRSRQRREFREKRLRGRKIS 285

Qy    451 CTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTA 510
          ||| ::: | |||| ::: :|| |::| ||| :| ||||||||:||
Db    286 PPSYARRDSPTYDPYKRSPSESSSESRSRSPTPGREEKITFITSFGGSDEEAAAAAAA 345

Qy    511 ACCAAGGGCTGGAGAGTGATGGAGCAGGAGGCCCACATGGAGGAGTGACAGCGTAGAC 570
          || :: :|||||||||||||: :|||:||||||||:: ::||::| :||:||||
Db    346 AAASGVTTGKPAPPQPGGPAPGRNASARRRSSSSSSASRTSSSRSSRSSSSRSRRG 405

Qy    571 ACAGAGGACAGCCGCTATGGAGCAGGAGGCTGAGGAGAGGCCCAGACGATGGAGTGC-- 628
          || ::|: ::|: | :::|:| ||| | ::::||||||: :|:| |||: :|| ::
Db    406 GGYYRSGRHARSRSRSWSRSRSRSRRYSRSRSRGRRHSGGGSRDGHRYSRSPARRGGYGP 465

Qy    629 -AGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCTCCAGGTAAACAGATTACAGAG 687
          ::: ::|::|: |::|| ::: :| :|| ||::::|:|| ::: :|:| ||::|:| :
Db    466 RRRSRSRSHSGDRYRRGGRGLRHHSSSRSSRSSWSLSPSRSRSLITRSRSHSPSQSRSRS 525

Qy    688 AAACTCAACTGCAAAGCCAACAGAAATATAGCCCAGAAAGGGAGATGG 747
          :::: :|::: |: :: :| :::|||||:||| :::|| ||:|::|:|
Db    526 RSRSQSPSPSPAREKLTRPAASPAVGEKLKKTEPAAGKETGAAKVTQADASGEAETEDAE 585

Qy    748 CAGCAGTGG 756
          |:|:|||||
Db    586 GAEQAVQGG 594

RESULT 15
US-10-141-759-10
; Sequence 10, Application US/10141759
; Publication No. US20030207361A1
; GENERAL INFORMATION:
;   APPLICANT: Baker,Kevin P.
```

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       May  4, 2004, 11:56:32 ; Search time 495 Seconds
                                        (without alignments)
                                        9835.222 Million cell updates/sec

Title:        US-10-644-659A-1
Perfect score: 1146
Sequence:     1 atggctccggcgaaaagga.........tgattacgctactcaagtga 1146

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:     3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

                            SUMMARIES

Database :     N_Geneseq_29Jan04:*
                1: geneseqn1980s:*
                2: geneseqn1990s:*
                3: geneseqn2000s:*
                4: geneseqn2001as:*
                5: geneseqn2001bs:*
                6: geneseqn2002s:*
                7: geneseqn2003as:*
                8: geneseqn2003bs:*
                9: geneseqn2003cs:*
               10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1144.4	99.9	1322	9	ADB63045	Adb63045 Human CDN
2	948.4	82.8	1373	5	AAS92182	Aas92182 DNA encod
c 3	499.4	43.6	537	4	AAI17811	Aai17811 Probe #77
c 4	499.4	43.6	537	4	ABA62763	Aba62763 Human foe
c 5	499.4	43.6	537	4	AAI42779	Aai42779 Probe #11
c 6	499.4	43.6	537	4	ABA30059	Aba30059 Probe #85
c 7	499.4	43.6	537	4	AAK36965	Aak36965 Human bon
c 8	499.4	43.6	537	4	AAK11153	Aak11153 Human bra
c 9	499.4	43.6	537	4	ABS36644	Abs36644 Human liv
c 10	363.4	31.7	365	4	AAI27007	Aai27007 Probe #16
c 11	363.4	31.7	365	4	ABA75275	Aba75275 Human foe
c 12	363.4	31.7	365	4	AAI55838	Aai55838 Probe #24
c 13	363.4	31.7	365	4	ABA39931	Aba39931 Probe #18
c 14	363.4	31.7	365	4	AAK49906	Aak49906 Human bon
c 15	363.4	31.7	365	4	AAK23820	Aak23820 Human bra
c 16	363.4	31.7	365	4	ABS49550	Abs49550 Human liv
17	205	17.9	261	3	AAC00343	Aac00343 Human sec
18	205	17.9	265	2	AAX40341	Aax40341 Human sec
19	56.8	5.0	10732	3	AAA10594	Aaa10594 Gene enco
20	55	4.8	552	4	ABL04119	Abl04119 Drosophil
21	55	4.8	2610	3	ABL04118	Abl04118 Drosophil
22	52.2	4.6	2000	7	ADA71938	Ada71938 Rice gene
23	49.8	4.3	1649	4	ABL06465	Abl06465 Drosophil

24	49.8	4.3	7090	4	ABL06464	Abl06464 Drosophil
25	42.4	3.7	2896	7	ACC47372	Acc47372 Human lat
26	42.4	3.7	3054	7	ACC47371	Acc47371 Rat late
27	40.4	3.5	14800	6	ABL66291	Abl66291 Lung canc
c 28	39.2	3.4	2000	7	ADA71938	Ada71938 Rice gene
29	39	3.4	6741	3	AAA10595	Aaa10595 Gene enco
30	38.8	3.4	14835	9	AAS94858	Aas94858 Human DNA
31	38.8	3.4	28564	9	ADD47107	Add47107 Rat gene
32	38.8	3.4	28564	9	ADD47113	Add47113 Rat gene
33	38.8	3.4	28564	9	ADD47117	Add47117 Rat gene
34	38.8	3.4	28564	9	ADD47111	Add47111 Rat gene
35	38.8	3.4	28564	9	ABB3334	Ade83334 Rat gene
36	38.8	3.4	28564	9	ADE57550	Ade57550 Rat gene
37	38.8	3.4	28564	9	ADE57554	Ade57554 Rat gene
c 38	38.6	3.4	6940	6	ABS73196	Abs73196 DNA encod
c 39	38.6	3.4	11910	5	ADC27511	Adc27511 Human MLL
c 40	38.6	3.4	12042	5	AAS76999	Aas76999 DNA encod
c 41	38.6	3.4	14255	2	AAQ43526	Aaq43526 ALL-1 gen
c 42	38.6	3.4	14255	2	AAQ75181	Aaq75181 ALL-1 (ac
43	38.6	3.4	14255	2	AAV20477	Aav20477 Human ALL
c 44	38.6	3.4	14255	2	ABL67248	Abl67248 Thyroid c
c 45	38.6	3.4	14255	6	ABS73200	Abs73200 Human tra

```
                            ALIGNMENTS

RESULT 1
ADB63045
ID   ADB63045 standard; cDNA; 1322 BP.
XX
AC   ADB63045;
XX
DT   04-DEC-2003 (first entry)
XX
DE   Human cDNA encoding clone SKMUS20026340.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH                   Location/Qualifiers
FT   Key
FT   CDS             57..1202
FT                   /tag= a
FT                   /product= "Clone SKMUS20026340 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI   Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI   Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB65015.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
```

XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 1322 BP; 399 A; 318 C; 372 G; 233 T; 0 U; 0 Other;

Query Match   99.9%; Score 1144.4; DB 9; Length 1322;
Best Local Similarity   99.9%;
Matches 1145; Conservative   0; Mismatches   0; Indels   1; Gaps   0;

Qy   1 ATGGCTCCGGGCGAAAAGGAAAAGCGGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATA 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   57 ATGGCTCCGGGCGAAAAGGAAAAGCGGGGAGGGCCCAGCCAAGAGCGCCCTCCGGAAGATA 116

Qy   61 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 CGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCGAATGAGAAC 176

Qy   121 AGCATCAGGCAGGCCCAGAGCCTACAGGCCTGGCTGCCCGGAGGGACCCAGGACTCACCT 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 AGCATCAGGCAGGCCCAGAGCCTACAGGCCTGGCTGCCCGGAGGGACCCAGGACTCACCT 236

Qy   181 CAAGGTCCTAAACCAATCACACCCCCTACTTCACACCCCTCAGAAGCTCAGAGTGCCCAAAG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CAAGGTCCTAAACCAATCACACCCCCTACTTCACACCCCTCAGAAGCTCAGAGTGCCCAAAG 296

Qy   241 TCGCCACCCCTGCCAGAAGGACATGGAAGGACATGAGATGGACAAAGCTCAGAAAGCCCCTGAG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 TCGCCACCCCTGCCAGAAGGACATGGAAGGACATGAGATGGACAAAGCTCAGAAAGCCCCTGAG 356

Qy   301 GTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 GTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAAACGGTGGTCAGCAAGACTTACGAGAGA 416

Qy   361 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGTGCTTGAACCT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 GGAGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGTGTGCTTGAACCT 476

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACCGCGGAGG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCCCAACCGCGGAGG 536

Qy   481 AGAAAATGTGCCAACCTGGTCTGCTGAGCTAACCAAGGGCTGGAGACTGATGGAGCAGGAG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 AGAAAATGTGCCAACCTGGTCTGCTGAGCTAACCAAGGGCTGGAGACTGATGGAGCAGGAG 596

Qy   541 GAGCCCACATGGAGGAGTGACAGCGTAGCACAGAGACACAGAGGCTATGGAGGAGGAGGCT 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   597 GAGCCCACATGGAGGAGTGACAGCGTAGCACAGAGACACAGAGGCTATGGAGGAGGAGGCT 656

Qy   601 GAGGAGAGGCCCGAGCAGGATGGAGTGGCGTGGTCAGGATCAAGCGCGCCCCTTG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   657 GAGGAGAGGCCCGAGCAGGATGGAGTGGCGTGGTCAGGATCAAGCGCGCCCCTTG 716

Qy   661 CCCTCCCAGTGAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAGAATATAGC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   717 CCCTCCCAGTGAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAGAATATAGC 776

Qy   721 CCAGTGGCAACTTGAAAGGGGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   777 CCAGTGGCAACTTGAAAGGGGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAG 836

Qy   781 AAGCTCAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   837 AAGCTCAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACAC 896

Qy   841 AAAGGAGGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCCAAG 900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   897 AAAGGAGGATGAGGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAGGGCCAAG 956

Qy   901 CGTGCTGAGGAGCACATCTACAGGGAAATGATGAGCATTCTCATTATCTGCACAATG 960
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   957 CGTGCTGAGGAGCACATCTACAGGGAAATGATGAGCATTCTCATTATCTGCACAATG 1016

Qy   961 GCTCGCCACAGACGACGGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1017 GCTCGCCACAGACGACGGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC 1076

Qy   1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAAACATGGACTGGTA 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1077 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAAACATGGACTGGTA 1136

Qy   1081 GACTTTGAAGAAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1137 GACTTTGAAGAAGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACGCTACTC 1196

Qy   1141 AAGTGA 1146
            ||||||
Db   1197 AAGTGA 1202

RESULT 2
AAS92182
ID   AAS92182 standard; cDNA; 1373 BP.
XX
AC   AAS92182;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27986.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG27995.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess

PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27986; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1373 BP; 396 A; 330 C; 408 G; 239 T; 0 U; 0 Other;

Query Match   82.8%; Score 948.4; DB 5; Length 1373;
Best Local Similarity   86.0%; Pred. No. 1e-259;
Matches 1145; Conservative   0; Mismatches   1; Indels 186; Gaps   1;

Qy   1 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGGCCAGCCAAGAGCGGCCCTCCGGAAGATA 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42 ATGGCTCCGGGCGAAAAGGAAAGCGGGAGGGGCCAGCCAAGAGCGGCCCTCCGGAAGATA 101

Qy   61 CGGCACAGCCACCTGGTCATCAGCTTGGCCCGAGGTTGGCGAGCAGTGGGCGAATGAGAAC 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 CGGCACAGCCACCCGGTCATCAGCTTGGCCCGAGGTTGGCGAGCAGTGGGCGGAATGAGAAC 161

Qy   121 AGCATCAGGCAGCCCAGGAGCTACAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCT 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   162 AGCATCAGGCAGCCCAGGAGCTACAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCT 221

Qy   181 CAAGCTCCTAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAGAGTGCCCCAAAG 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   222 CAAGCTCCTAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAGAGTGCCCCAAAG 281

Qy   241 TCGGCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAG 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   282 TCGGCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAG 341

Qy   301 GTTTCTCACATCAAAAGAAGAGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGA 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   342 GTTTCTCACATCAAAAGAAGAGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGA 401

Qy   361 GGAGGGACGTGAGCCACCTCAGCCACCAGGTACGAGAGGGATGCTGGTGTGCTTGAACCT 420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   402 GGAGGGACGTGAGCCACCTCAGCCACCAGTGTACGAGAGGGATGCTGGTGTGCTTGAACCT 461

Qy   421 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCCCCAACGCGGAGG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   462 GGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCCCCAACGCGGAGG 521

Qy   481 AGAAAATGTGCCAACCTGGTGCTCGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   522 AGAAAATGTGCCAACCTGGTGCTCGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAG 581

Qy   541 GAGCCCACCATGGAGGAGTGACAGCAGAGACAGACAGCAGCCGCTATGGAGGAGGAGGCT 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   582 GAGCCCACCATGGAGGAGTGACAGCAGAGACAGACAGCAGCCGCTATGGAGGAGGAGGCT 641

Qy   601 GAGGAGAGGGCCCGAGCAGGATGGAGTGCAGGTGCGTGTCGTCGTCAGGATCGAGGATCAAGCGCCCTTG 660

Db   642 GAGGAGAGGGCCCGAGCAGCAGGATGGAGTGCAGGTGGGCTGTGGTCAGGATCAAGCGCGCCCCTTG 701
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   661 CCCTCCCA---------------------- 668
           |||||||
Db   702 CCCTCCCCAAGGAGGGGTGCTGCGATGGGGTGCTGTCGATGAGGAGGAAATACTCCCTCTACAGGAAGATCTAT 761

Qy   669 ---------------------- 668
Db   762 GGAGTTGGCAGTGAAAGCCCATTTCCACTTTCTCCATACAGATCGGAGACAGAAGAGAGAAA 821

Qy   669 ---------------------- 668
Db   822 ATCATGCCTTTAGGCGTCCCATCCATTTGAAGCTGCCTTTGGTAACCAGATTTCCGGA 881

Qy   669 ------------GGTAAACAGATTTACAGAGAAGAAACTCAACTGCAAAGCCCAACAGAAA 714
              |||||||||||||||||||||||||||||||||||||||
Db   882 AGTAGGTTGGATGGGTAAACAGATTTACAGAGAAGAAACTCAACTGCAAAGCCCAACAGAAA 941

Qy   715 TATAGCCCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACACATACAA 774
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   942 TATAGCCCCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGCTGATGAACACATACAA 1001

Qy   775 TCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGC 834
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1002 TCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTGATTACGAGCTGGCCATGTCCACCCGC 1061

Qy   835 CTACCAAAGGAGATGAGGGCTATGGGCCGCCCAAAGAAGGAACCAAAACTGCTGAAAGG 894
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1062 CTACCAAAGGAGATGAGGGCTATGGGCCGCCCAAAGAAGGAACCAAAACTGCTGAAAGG 1121

Qy   895 GCCAAGCCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGC 954
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1122 GCCAAGCCGTGCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGC 1181

Qy   955 ACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGAC 1014
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1182 ACAATGGCTCGCCACAGACGAGATGGCAAGATCCAGGTTACTTTTGGAGATCTCTTTGAC 1241

Qy   1015 AGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGA 1074
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1242 AGATACGTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGA 1301

Qy   1075 CTGGTAGACTTTGAAGGAGGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACG 1134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1302 CTGGTAGACTTTGAAGGAGGAGATGCTATGGCAAGGCCGAGATGACCATGTTGTGATTACG 1361

Qy   1135 CTACTCAAGTGA 1146
           |||||||||||
Db   1362 CTACTCAAGTGA 1373

RESULT 3
AAI17811/c

RESULT 4
ABA62763/c
ID   ABA62763 standard; DNA; 537 BP.
XX
AC   ABA62763;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #11068.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 11068; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match   43.6%; Score 499.4; DB 4; Length 537;
Best Local Similarity   99.8%; Pred. No. 7e-132;
Matches 500; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   169 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAG 228
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACCAGAAAGCTCAG 478

Qy   229 AGTGCCCAAAGTCGCACCCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAG 288
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGTGCCCAAAGTCGCACCCCGCCTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAG 418

Qy   289 AAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAG 348
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AAAGCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGGTCAGCAAG 358

Qy   349 ACTTACGAGAGAGGAGGGGACCTGAGCCACCTGAGCCACGGGGATGCTGGT 408
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 ACTTATGAGAGAGGAGGGGACGTGAGCCACCTGAGCCACGGGGATGCTGGT 357

Qy   409 GTGCTTGAACCTGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 238

Qy   649 AAGCGCCCCTTGCCCTCCAG 669
          ||||||||||||||||||||
Db   57 AAGCGCCCCTTGCCCTCCAG 37

```
Qy  229 AGTGCCCAAAGTCGCCACCCCGCCTGCCAGGAAGGACATGGAGATGGACAAAGCTTCAGAG 288
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  477 AGTGCCCAAAGTCGCCACCCCGCCTGCCAGGAAGGACATGGAGATGGACAAAGCTTCAGAG 418

Qy  289 AAAGCCCTGAGGTTTCCACATCAAAAGAAAGAGAGGTGTCAAAACGTGGTCAGCAAG 348
        ||||||||||||||||||||||||||||||||||| ||||||||| |||||||||||
Db  417 AAAGCCCTGAGGTTTCCACATCAAAAGAAAGAAGAGGTGTCAAAACGGTGGTCAGCAAG 358

Qy  349 ACTTACGAGAGAGAGGGGACGTGAGCCACCTCAGCCACAGTACGAGAGGATGGTGGT 408
        ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  357 ACTTATGAGAGAGAGGGGACGTGAGCCACCTCAGCCACAGTACGAGAGGATGCTGGT 298

Qy  409 GTGCTTGAACCTGGCAGCAGCAGAGAATGACATTGACAGAATCCTCCACAGCCACGCGGTCC 468
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  297 GTGCTTGAACCTGGCAGCAGCAGAGAATGACATTGACAGAATCCTCCACAGCCACGCGGTCC 238

Qy  469 CCAACGCGGAGGAGGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGGCTGGAGAGTG 528
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  237 CCAACGCGGAGGAGGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGGCTGGAGAGTG 178

Qy  529 ATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTAT 588
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  177 ATGGAGCAGGAGGAGCCCACATGGAGGAGTGACAGCGTAGACACAGAGGACAGCGGCTAT 118

Qy  589 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGTGTGGTCAGGATC 648
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  117 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGATGGAGTGCAGGTGGCTGTGTGGTCAGGATC 58

Qy  649 AAGCGCCCCTTGCCCTCCCAG 669
        |||||||||||||||||||||
Db   57 AAGCGCCCCTTGCCCTCCCAG 37
```

RESULT 5
AAI42779/c
ID   AAI42779 standard; DNA; 537 BP.
XX
AC   AAI42779;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #11465 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human placenta.
XX
PS   Claim 25; SEQ ID NO 11465; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders
XX
SQ   Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

```
Query Match      43.6%;  Score 493.4;  DB 4;  Length 537;
Best Local Similarity    99.8%;  Pred. No. 7e-132;
Matches    500;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy  169 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACCACCAGAAAGCTCAG 228
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  537 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACCACCAGAAAGCTCAG 478
```

RESULT 6
ABA30059/c
ID   ABA30059 standard; DNA; 537 BP.
XX
AC   ABA30059;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Probe #8525 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000666.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488899/53.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts.
XX

```
PS   Claim 1; SEQ ID NO 8525; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease. Note:The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match        43.6%;   Score 499.4;  DB 4;   Length 537;
Best Local Similarity   99.8%;   Pred. No. 7e-132;
Matches  500;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   169 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 478

Qy   229 AGTGCCCCAAAGTCGCCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGTGCCCCAAAGTCGCCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 418

Qy   289 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACCGTGGTCAGCAAG 348
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACCGTGGTCAGCAAG 358

Qy   349 ACTTACGAGAGAGGAGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGT 408
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 ACTTATGAGAGAGGAGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGT 298

Qy   409 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 468
         |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   297 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 238

Qy   469 CCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTG 528
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CCAACGCGGAGGAGAAAATGTGCCAACCTCGTGTCTGAGCTAACCAAGGGCTGAGAGTG 178

Qy   529 ATGGAGCAGGAGGAGCCCACATGGAGGAGTTGACAGCGTAGACACAGAGACAGCAGGGCTAT 588
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 ATGGAGCAGGAGGAGCCCACATGGAGGAGTTGACAGCGTAGACACAGAGACAGCAGGGCTAT 118

Qy   589 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGAGTGCAGGTGGCTGTGGTCAGGATC 648
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGAGTGCAGGTGGCTGTGGTCAGGATC 58

Qy   649 AAGCGGCCCCTTGCCCTCCCAG 669
         ||||||||||||||||||||||
Db    57 AAGCGGCCCCTTGCCCTCCCAG 37
```

```
RESULT 7
AAK36965/c
ID   AAK36965 standard; DNA; 537 BP.
XX
AC   AAK36965;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed single exon probe SEQ ID NO: 11522.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
```

```
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human bone marrow.
XX
PS   Example 4; SEQ ID NO 11522; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention
XX
SQ   Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match        43.6%;   Score 499.4;  DB 4;   Length 537;
Best Local Similarity   99.8%;   Pred. No. 7e-132;
Matches  500;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   169 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 478

Qy   229 AGTGCCCCAAAGTCGCCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGTGCCCCAAAGTCGCCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 418

Qy   289 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACCGTGGTCAGCAAG 348
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACCGTGGTCAGCAAG 358

Qy   349 ACTTACGAGAGAGGAGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGT 408
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 ACTTATGAGAGAGGAGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGGATGCTGGT 298

Qy   409 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 468
         |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   297 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 238

Qy   469 CCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTG 528
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CCAACGCGGAGGAGAAAATGTGCCAACCTCGTGTCTGAGCTAACCAAGGGCTGAGAGTG 178

Qy   529 ATGGAGCAGGAGGAGCCCACATGGAGGAGTTGACAGCGTAGACACAGAGACAGCAGGGCTAT 588
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 ATGGAGCAGGAGGAGCCCACATGGAGGAGTTGACAGCGTAGACACAGAGACAGCAGGGCTAT 118

Qy   589 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGAGTGCAGGTGGCTGTGGTCAGGATC 648
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 GGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGAGTGCAGGTGGCTGTGGTCAGGATC 58

Qy   649 AAGCGGCCCCTTGCCCTCCCAG 669
         ||||||||||||||||||||||
Db    57 AAGCGGCCCCTTGCCCTCCCAG 37
```

RESULT 8
AAK11153/c
ID AAK11153 standard; DNA; 537 BP.
XX
AC AAK11153;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 11144.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
PN WO20157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
XX
PS Example 4; SEQ ID NO 11144; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match   43.6%; Score 499.4; DB 4; Length 537;
Best Local Similarity 99.8%; Pred. No. 7e-132;
Matches 500; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   169 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCTACTTCACACCAGAAAGCTCAG 228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CAGGACTCACCTCAAGTCCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 478

Qy   229 AGTGCCCAAAGTCGCCACCCCCGCCTGCCAGGAGGACATGGAGATGGACAAAGCTCAGAG 288
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGTGCCCCAAAGTCGCCACCCCCGCCTGCCAGGAGGACATGGAGATGGACAAAGCTCAGAG 418

Qy   289 AAAGCCCCTGAGGTTTCCACATCAAAAGAAAAGAGGTGTCCAAAACGTTGGTCAGCAAG 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AAAGCCCCTGAGGTTTCCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAAG 358

Qy   349 ACTTACGAGAGGAGGGGACGTGAGGCACTCAGCCACCAGTACGAGAGGGGATGCTGGT 408
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 ACTTATGAGAGGAGGGGGACGTGAGGCACTCAGCCACCAGTACGAGAGGGGATGCTGGT 298

Qy   409 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCC 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACCGGCTCC 238

Qy   469 CCAACGCGGAGAGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGAGAGTG 528
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CCAACGCGGAGAGAAAATGTGCCAACCTGGTCTGAGCTAACCAAGGGCTGAGAGTG 178

Qy   529 ATGGAGCAGGAGGAGCCCACATGGAGGTAGACACGGTAGACACAGAGGACAGCGGCTAT 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 ATGGAGCAGGAGGAGCCCACATGGAGGTAGACACGGTAGACACAGAGGACAGCGGCTAT 118

Qy   589 GGAGGAGGAGGCTGAGGAGGCCCAGACAGCAGGATGAGTGCGTGTGGTCAGGATC 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 GGAGGAGGAGGCTGAGGAGGCCCAGACAGCAGGATGGAGTGCGTGTGGTCAGGATC 58

Qy   649 AAGCGCCCTTGCCCTCCCAG 669
           ||||||||||||||||||||
Db   57 AAGCGCCCTTGCCCTCCCAG 37

RESULT 9
ABS36644/c
ID ABS36644 standard; DNA; 537 BP.
XX
AC ABS36644;
XX
DT 25-FEB-2003 (first entry)
XX
DE Human liver single exon probe, SEQ ID No 11634.
XX
KW Human; single exon nucleic acid probe; liver; cirrhosis;
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW coronary heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO20157273-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 1; SEQ ID NO 11634; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed

CC   specification but was obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 537 BP; 82 A; 161 C; 142 G; 152 T; 0 U; 0 Other;

Query Match   43.6%;   Score 499.4; DB 4;   Length 537;
Best Local Similarity   99.8%;   Pred. No. 7e-132;
Matches 500; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   169 CAGGACTCACCTCAAGTCTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   537 CAGGACTCACCTCAAGTCTCTAAACCAATCACACCCCCTACTTCACACCAGAAAGCTCAG 478

Qy   229 AGTGCCCAAAGTCGGCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   477 AGTGCCCAAAGTCGGCACCCGCCTGCCGAAGGACATGGAGATGGACAAAGCTCAGAG 418

Qy   289 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAAG 348
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   417 AAAGCCCCTGAGGTTTCTCACATCAAAAGAAAAGAGGTGTCCAAAACGGTGGTCAGCAAG 358

Qy   349 ACTTACGAGAGAGGGGGACGTGAGCCACCTCAGCCACGGTACGAGAGGGATGCTGGT 408
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   357 ACTTATGAGAGAGGGGGACGTGAGCCACCTCAGCCACGGTACGAGAGGGATGCTGGT 298

Qy   409 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 468
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   297 GTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAGCCACGGCTCC 238

Qy   469 CCAACGCGGAGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTG 528
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   237 CCAACGCGGAGAGGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGAGAGTG 178

Qy   529 ATGGAGAGGAGGAGCCACATGGAGGAGTGACACAGAGGACAGAGGACAGCGGCTAT 588
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 ATGGAGGAGGAGGAGCCACATGGAGGAGTGACACAGAGGACAGAGGACAGCGGCTAT 118

Qy   589 GGAGGAGAGGCTGAGGAGAGCCCGAGCACGGATGAGTGCAGGTGGCTGTGGTCAGGATC 648
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   117 GGAGGAGAGGCTGAGGAGAGCCCGAGCACGGATGAGTGCAGGTGGCTGTGGTCAGGATC 58

Qy   649 AAGCGCCCCTTGCCCTCCCAG 669
         |||||||||||||||||||||
Db   57 AAGCGCCCCTTGCCCTCCCAG 37

RESULT 10
AAI27007/c
ID   AAI27007 standard; DNA; 365 BP.
XX
AC   AAI27007;
XX
DT   12-OCT-2001 (first entry)
XX
DE   Probe #16940 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015727B-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00612366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.

PR   04-OCT-2000; 2000GB-0002426J.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 16940; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

Query Match   31.7%;   Score 363.4; DB 4;   Length 365;
Best Local Similarity   99.7%;   Pred. No. 3.5e-93;
Matches 364; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   260 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAGA 319
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAGA 306

Qy   320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGGGACGTGAGCCACC 379
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGAGGGGGACGTGAGCCACC 246

Qy   380 TCAGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACA 439
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 TCAGCCACAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACA 186

Qy   440 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGAGGAGAAAATGTGCCAACCTGG 499
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGAGGAGAAAATGTGCCAACCTGG 126

Qy   500 TGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGGAGCCACCATGGAGGAGTG 559
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 TGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGGAGCCACCATGGAGGAGTG 66

Qy   560 ACAGGCTAGACACAGAGGACAGCCGCTATGGAGGAGGCTGAGGAGAGCCCGAGCAGG 619
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 ACAGGCTAGACACAGAGGACAGCCGCTATGGAGGAGGCTGAGGAGAGCCCGAGCAGG 6

Qy   620 ATGGA 624
         |||||
Db   5 ATGGA 1

RESULT 11
ABA75275/c
ID   ABA75275 standard; DNA; 365 BP.
XX
AC   ABA75275;
XX
DT   01-FEB-2002 (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #23580.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX

PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00612366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
     gene expression in human fetal liver.
XX
PS   Claim 4; SEQ ID NO 23580; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

Query Match   31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity   99.7%; Pred. No. 3.5e-93;
Matches 364;   Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   260 AAGGACAATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 AAGGACAATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 306

Qy   320 AAGAGGTGTCCAAAACGGTGTCAGCAAGACTTACGAGAGAGGGGACGTGAGCCACC 379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 AAGAGGTGTCCAAAACGGTGTCAGCAAGACTTATGAGAGAGGGGACGTGAGCCACC 246

Qy   380 TCAGCCACAGGTACGAGAGGGATGCTGGTCGTGCTTGAACCTGGCAGCCAGAGAATGACA 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 TCAGCCACAGGTACGAGAGGGATGCTGGTCGTGCTTGAACCTGGCAGCCAGAGAATGACA 186

Qy   440 TTGACAGAATCCTCCACAGCCACGGCTCCCACCGGAGGAGAAAATGTGCCAACCTGG 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 TTGACAGAATCCTCCACAGCCACGGCTCCCACCGGAGGAGAAAATGTGCCAACCTGG 126

Qy   500 TGTCTGAGCTAACCAAGGCTGGAGAGTGATGGAGCGCCCACCATGGAGGAGGTG 559
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 TGTCTGAGCTAACCAAGGCTGGAGAGTGATGGAGCGCCCACCATGGAGGAGGTG 66

Qy   560 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGACAGCGGCTGAGGAGGCCCGAGCAGG 619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGACAGCGGCTGAGGAGGCCCGAGCAGG 6

Qy   620 ATGGA 624
          |||||
Db   5 ATGGA 1

RESULT 12
AAI55838/c
ID   AAI55838 standard; DNA; 365 BP.
XX

AC   AAI55838;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #24524 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
     genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000663.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00612366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
     gene expression in human placenta.
XX
PS   Claim 25; SEQ ID NO 24524; 654pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

Query Match   31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity   99.7%; Pred. No. 3.5e-93;
Matches 364;   Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   260 AAGGACAATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 AAGGACAATGGAGATGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 306

Qy   320 AAGAGGTGTCCAAAACGGTGTCAGCAAGACTTACGAGAGAGGGGACGTGAGCCACC 379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 AAGAGGTGTCCAAAACGGTGTCAGCAAGACTTATGAGAGAGGGGACGTGAGCCACC 246

Qy   380 TCAGCCACAGGTACGAGAGGGATGCTGGTCGTGCTTGAACCTGGCAGCCAGAGAATGACA 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   245 TCAGCCACAGGTACGAGAGGGATGCTGGTCGTGCTTGAACCTGGCAGCCAGAGAATGACA 186

Qy   440 TTGACAGAATCCTCCACAGCCACGGCTCCCACCGGAGGAGAAAATGTGCCAACCTGG 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 TTGACAGAATCCTCCACAGCCACGGCTCCCACCGGAGGAGAAAATGTGCCAACCTGG 126

Qy   500 TGTCTGAGCTAACCAAGGCTGGAGAGTGATGGAGCGCCCACCATGGAGGAGGTG 559
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 TGTCTGAGCTAACCAAGGCTGGAGAGTGATGGAGCGCCCACCATGGAGGAGGTG 66

Qy   560 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGACAGCGGCTGAGGAGGCCCGAGCAGG 619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 ACAGCGTAGACACAGAGGACAGCGGCTATGGAGGACAGCGGCTGAGGAGGCCCGAGCAGG 6

Qy   620 ATGGA 624

Db   5 ATGGA 1
            |||||

RESULT 13
ABA39931/c
ID   ABA39931 standard; DNA; 365 BP.
XX
AC   ABA39931;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Probe #18397 for gene expression analysis in human heart cell sample.
XX
KW   Human; gene expression; heart; microarray; vascular system; probe;
KW   cardiovascular disease; hypertension; cardiac arrhythmia;
KW   congenital heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015274-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000666.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   hearts.
XX
PS   Claim 4; SEQ ID NO 18397; 530pp; English.
XX
CC   The present invention relates to single exon nucleic acid probes for
CC   measuring human gene expression in a sample derived from human heart. The
CC   present sequence is one such probe. The probes may be used for
CC   predicting, measuring and displaying gene expression in samples derived
CC   from the human heart via microarrays. By measuring gene expression, the
CC   probes are useful for predicting, diagnosing, grading, staging,
CC   monitoring and prognosing diseases of the human heart and vascular system
CC   e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC   congenital heart disease. Note: The sequence data for this patent did not
CC   form part of the printed specification, but was obtained in electronic
CC   format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

Query Match   31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity   99.7%; Pred. No. 3.5e-93;
Matches 364; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   260 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 319
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 306

Qy   320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGAGGGACGTGAGCCACC 379
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGAGGGACGTGAGCCACC 246

Qy   380 TCAGCCACAGGTACGAGAGGGATGCTGGTGTCGTTGAACCTGGCAGCCAGAGAATGACA 439
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   245 TCAGCCACAGGTACGAGAGGGATGCTGGTGTCGTTGAACCTGGCAGCCAGAGAATGACA 186

Qy   440 TTGACAGAATCCTCCACAGCCCACGCGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTG 499
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 TTGACAGAATCCTCCACAGCCCACGCGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTG 126

Qy   500 TGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCCCACATGGAGGAGTG 559
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 TGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCCCACATGGAGGAGTG 66

Qy   560 ACAGGCGTAGACACAGAGGACACAGCGGCTATGGAGGAGGAGGGTGAGGAGAGCCCGAGCAGG 619
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 ACAGGCGTAGACACAGAGGACACAGCGGGCTATGGAGGAGGAGGGTGAGGAGAGCGCGCCGAGCAGG 6

Qy   620 ATGGA 624
          |||||
Db   5 ATGGA 1

RESULT 14
AAK49906/c
ID   AAK49906 standard; DNA; 365 BP.
XX
AC   AAK49906;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human; bone marrow expressed single exon probe SEQ ID NO: 24463.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000668.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-0060840B.
PR   03-AUG-2000; 2000US-0063236G.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236355P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human bone marrow.
XX
PS   Example 4; SEQ ID NO 24463; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC   the probes of the invention
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

Query Match   31.7%; Score 363.4; DB 4; Length 365;
Best Local Similarity   99.7%; Pred. No. 3.5e-93;
Matches 364; Conservative   0; Mismatches   1; Indels   1; Gaps   0;

Qy   260 AAGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGA 319

CC   diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC   epilepsy and cancers. The present sequence is one of the probes of the
CC   invention
XX
SQ   Sequence 365 BP; 49 A; 126 C; 78 G; 112 T; 0 U; 0 Other;

```
Query Match             31.7%;  Score 363.4;  DB 4;  Length 365;
Best Local Similarity   99.7%;  Pred. No. 3.5e-93;
Matches  364;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    260 AAGGACATGGAGATGGACAAAGCTCAGAGAAGCCCCTGAGGTTTCTCACATCAAAAAGA 319
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    365 AAGGACATGGAGATGGACAAAGCTCAGAGAAGCCCCTGAGGTTTCTCACATCAAAAAGA 306

Qy    320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACGTGAGCCACC 379
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    305 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGGAGGGGACGTGAGCCACC 246

Qy    380 TCAGCCACAGGTACGAGAGGGGATGGCTCGTGTGCTTGAACCTGGCAGCCAGAGAATGACA 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 TCAGCCACAGGTACGAGAGGGGATGGCTCGTGTGCTTGAACCTGGCAGCCAGAGAATGACA 186

Qy    440 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGG 499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGG 126

Qy    500 TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCCAGGAGGAGAGGCTGAGGAGGAGTG 559
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCCAGGAGGAGAGGCTGAGGAGGAGTG 66

Qy    560 ACAGCGTAGCACAGAGGACAGCGGCTTATGGAGGAGAGGCTGAGGAGGGCCCGAGCAGG 619
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 ACAGCGTAGCACAGAGGACAGCGGCTTATGGAGGAGAGGCTGAGGAGGGCCCGAGCAGG 6

Qy    620 ATGGA 624
          |||||
Db      5 ATGGA 1
```

Search completed: May 5, 2004, 00:59:14
Job time : 503 secs

```
Db    365 AAGGACATGGAGATGGACAAAGCTCAGAGAAGCCCCTGAGGTTTCTCACATCAAAAAGA 306

Qy    320 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTACGAGAGGAGGGGACGTGAGCCACC 379
          |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    305 AAGAGGTGTCCAAAACGGTGGTCAGCAAGACTTATGAGAGGAGGGGACGTGAGCCACC 246

Qy    380 TCAGCCACAGGTACGAGAGGGGATGGCTCGTGTGCTTGAACCTGGCAGCCAGAGAATGACA 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    245 TCAGCCACAGGTACGAGAGGGGATGGCTCGTGTGCTTGAACCTGGCAGCCAGAGAATGACA 186

Qy    440 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGG 499
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    185 TTGACAGAATCCTCCACAGCCACGGCTCCCCAACGCGGAGGAGAAAATGTGCCAACCTGG 126

Qy    500 TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCCAGGAGGAGAGGCTGAGGAGGAGTG 559
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    125 TGTCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCCAGGAGGAGAGGCTGAGGAGGAGTG 66

Qy    560 ACAGCGTAGCACAGAGGACAGCGGCTTATGGAGGAGAGGCTGAGGAGGGCCCGAGCAGG 619
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     65 ACAGCGTAGCACAGAGGACAGCGGCTTATGGAGGAGAGGCTGAGGAGGGCCCGAGCAGG 6

Qy    620 ATGGA 624
          |||||
Db      5 ATGGA 1
```

RESULT 15
AAK23820/c
ID   AAK23820 standard; DNA; 365 BP.
XX
AC   AAK23820;
XX
DT   05-NOV-2001 (first entry)
XX
DE   Human brain expressed single exon probe SEQ ID NO: 23811.
XX
KW   Human; brain expressed exon; gene expression analysis; probe; microarray;
KW   Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW   ss.
XX
OS   Homo sapiens.
XX
PN   WO200157275-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000667.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0236687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-483446/52.
XX
PT   Single exon nucleic acid probes for analyzing gene expression in human
PT   brains.
XX
PS   Example 4; SEQ ID NO 23811; 650pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   brain. They can be used to measure gene expression in brain cell samples,
CC   which may enable the diagnosis and improved treatment of nervous system

```
                    GenCore version 5.1.6
               Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on:       May  4, 2004, 08:49:08 ; Search time 18 Seconds
                                       (without alignments)
                                       6630.266 Million cell updates/sec

Title:       US-10-644-659A-1
Perfect score:  2058
Sequence:    1 atggctccgggcgaaagga........tgattacgctactcaagtga 1146

Scoring table:  BLOSUM62
         Xgapop  10.0 , Xgapext  0.5
         Ygapop  10.0 , Ygapext  0.5
         Fgapop   6.0 , Fgapext  7.0
         Delop    6.0 , Delext   7.0

Searched:    141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:       283362

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPro_spool/p/US10644659/runat_04052004_084902_28510/app_query.fasta_1.1287
 -DB=SwissProt_42 -QMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US10644659 @CGN_1_1.16 @runat_04052004_084902_28510 -LONGLOG -ICPU=3
 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
 -DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   SwissProt_42:*

  Pred. No. is the number of results predicted by chance to have a
  score greater than or equal to the score of the result being printed,
  and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	130.5	6.3	1739	1	CHD2_HUMAN	O14647 homo sapien
2	123	6.0	838	1	GLT4_WHEAT	P08489 triticum ae
3	122	5.9	887	1	NED4_RAT	Q62940 rattus norv
4	121	5.9	2805	1	MAPA_HUMAN	P78559 homo sapien
5	120	5.8	839	1	GLT5_WHEAT	P10388 triticum ae
6	120	5.8	1182	1	HAIR_MOUSE	Q61645 mus musculu
7	119.5	5.8	458	1	YNE1_CAEEL	P30640 caenorhabdi
8	118.5	5.8	1067	1	MICA_HUMAN	Q8tdz2 homo sapien
9	116.5	5.7	1001	1	IF2_STNY3	P72689 synechocyst
10	115.5	5.6	2472	1	NCR2_MOUSE	Q9wu42 mus musculu
11	115	5.6	646	1	SG1_BOVIN	P23389 bos taurus
12	114	5.5	1275	1	TRP_DROME	P19334 drosophila
13	114	5.5	1411	1	TCOF_HUMAN	Q13428 homo sapien
14	113.5	5.5	1293	1	XPC_DROME	Q24595 drosophila
15	113	5.5	350	1	VSX1_CHICK	Q9ial2 gallus gall
16	113	5.5	1125	1	IF2_PROMM	Q7v5m4 prochloroco
17	113	5.5	1593	1	AT12_HUMAN	P58397 homo sapien
c 18	113	5.4	2003	1	NTC4_HUMAN	Q99466 homo sapien

```
                                           ALIGNMENTS

RESULT 1
CHD2_HUMAN
ID   CHD2_HUMAN     STANDARD;      PRT;  1739 AA.
AC   O14647;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Chromodomain-helicase-DNA-binding protein 2 (CHD-2).
GN   CHD2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97470991; PubMed=9326634;
RA   Woodage T., Basrai M.A., Baxevanis A.D., Hieter P., Collins F.S.;
RT   "Characterization of the CHD family of proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:11472-11477(1997).
-!- FUNCTION: Sequence-selective DNA-binding protein (By similarity).
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
-!- SIMILARITY: Belongs to the SNF2/RAD54 helicase family.
-!- SIMILARITY: Contains 2 chromo domains.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics  and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF006514; AAB87382.1; -.
DR   HSSP; P23197; IAP0.
DR   Genew; HGNC:1917; CHD2.
DR   MIM; 602119; -.
DR   GO; GO:0004003; F:ATP dependent DNA helicase activity; TAS.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   GO; GO:0007001; P:chromosome organization and biogenesis (sen. . .; TAS.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
DR   InterPro; IPR000953; Chromo.
DR   InterPro; IPR001410; DEAD.
```

19	112.5	5.5	255	1	HXD4_HUMAN	P09016 homo sapien
20	112.5	5.5	6632	1	UN89_CAEEL	O01761 caenorhabdi
21	112	5.4	728	1	TRDN_HUMAN	Q13061 homo sapien
22	111	5.4	677	1	SG1_HUMAN	P05060 homo sapien
23	110.5	5.4	671	1	VINE_HUMAN	O60504 homo sapien
24	110.5	5.4	1217	1	AF4_MOUSE	O88573 mus musculu
25	110	5.4	2774	1	MAPA_RAT	P34926 rattus norv
26	110	5.3	459	1	RN25_HUMAN	Q96bh1 homo sapien
27	109.5	5.3	675	1	SG1_RAT	O35314 rattus norv
28	109.5	5.3	1336	1	W146_HUMAN	Q9c0j8 homo sapien
29	109.5	5.3	1394	1	CNG4_BOVIN	Q28181 bos taurus
30	109	5.3	660	1	YHL1_EBV	P03181 epstein-bar
31	109	5.3	753	1	ELA2_HUMAN	Q8iyf1 homo sapien
32	109	5.3	3664	1	MINT_HUMAN	Q96658 homo sapien
33	108.5	5.3	486	1	MEC2_HUMAN	P51608 homo sapien
34	108.5	5.3	1181	1	HAIR_RAT	P97609 rattus norv
35	108.5	5.3	1448	1	TRC5_ECOLI	P27190 escherichia
36	108	5.2	317	1	RX2_CHICK	Q9pvx0 gallus gall
37	108	5.2	1211	1	SBCC_PSEAE	Q9hwb8 pseudomonas
38	108	5.2	3644	1	MINT_MOUSE	Q62504 mus musculu
39	107.5	5.2	363	1	VSX1_MOUSE	Q9iv10 mus musculu
40	107.5	5.2	486	1	AST8_DROME	P09775 drosophila
41	107.5	5.2	931	1	MASK_XENLA	Q9ptg8 xenopus lae
42	107.5	5.2	1324	1	IRS2_HUMAN	Q9y4h2 homo sapien
43	107.5	5.2	2339	1	CCAB_RABIT	Q05152 oryctolagus
44	107	5.2	1067	1	SGG_DROME	P18431 drosophila
45	107	5.2	1227	1	PR16_HUMAN	Q92620 homo sapien

```
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR000330; SNF2_N.
DR   Pfam; PF00385; chromo; 2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00176; SNF2_N; 1.
DR   SMART; SM00298; CHROMO; 2.
DR   SMART; SM00487; DEXDc; 1.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00598; CHROMO_1; 2.
DR   PROSITE; PS50013; CHROMO_2; 2.
DR   PROSITE; PS00690; DEAH_ATP_HELICASE; FALSE_NEG.
KW   Hydrolase; Helicase; Nuclear protein; ATP-Binding; DNA-binding;
KW   Repeat.
FT   DOMAIN        1     74       SER-RICH.
FT   DOMAIN      121    139       SER-RICH.
FT   DOMAIN      261    353       CHROMO 1.
FT   DOMAIN      378    456       CHROMO 2.
FT   NP_BIND     509    516       ATP (POTENTIAL).
FT   SITE        617    620       DEAH BOX.
SQ   SEQUENCE   1739 AA;  200560 MW;  D62089C25EBAA40F CRC64;

Alignment Scores:
Pred. No.:                0.163
Score:                  130.50
Percent Similarity:      37.10%
Best Local Similarity:   22.26%
Query Match:              6.34%
DB:                           1

US-10-644-659A-1 (1-1146) x CHD2_HUMAN (1-1739)
```

```
Qy  109 GCGAATGAGAACAGCATCAGGCAGGCCCAGAGCCTACAGGCTGGCTGCCGGGAGGGAGGAGGACC 168
            ::::||| :|:||:    |:|:::||  :::::|||  :|::|:|||
Db   17 SerAsnAlaSerSerHisSerAlaSerGluGlnAlaSerGlyGluAlaSerGlySerAspSerGlySerGln  36

Qy  169 CAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCACAC--CAGAAAGCT 225
            ::|||| |||          |||  ::|  ||| :|:||:
Db   37 SerGluSerGluGlnGlySer------AspProGlySerGlyHisGlySerGluSer  53

Qy  226 CAGAGCGCCCAAAGTCGCCACCCCGCTCGCCAGAAGGACATGGAGGATGGACAAAGCTCA 285
            |||||||    ::|||        :::    ||
Db   54 AsnSerSerGluSerGluSerGlnSerGluSerGluSerGluSerAlaGlySer  73

Qy  286 GAGAAAGCCCCTGAGGTTTCCACATCAAAAAGAAAGAGGTGTCCAAAACGGTGTCAGC 345
            :::  :|            ||||:||                |||||||
Db   74 LysSerGlnProValleuProGluAlaLysGluLysProAlaSerLysLys---------  90

Qy  346 AAGACTTACGAGAGGAGGAGGGGACTGAGCCACCTCAGCCACAGGTACGAGAGGGATGCT 405
                                  :::                     |||
Db   91 -------GluArgIleAlaAspValLysMetTrpGluGluTyrProAspValTyr 107

Qy  406 GGTGTGCTTGAACCTGGCAGCCAGAGAATGACATTGACAGAATCCTC-------------- 453
            ||||||          |||     ::|        |||  :::
Db  108 GlyValArgArgSerAsnArgSerArgGlnGluProSerArgPheAsnIleLysGluGlu 127

Qy  454 ---------CACAGCACGGCTCCCCAACGCGGAGGAGAAAAATGTGCCAACCTGGTG 501
                       ||||||||
Db  128 AlaSerSerGlySerGluSerGlyGlySerProLysSerArgArg----------- 140

Qy  502 TCTGAGCTAACCAAGGGCTGAGAGTGATGGAGCAGGAGGAGGCCACCATGGAGGAGTGAC 561
                      ||||||||||                 :::
Db  141 -------------GlyGlnArgGlnLeuLysLysGlnGluLysTrpLysGlnGluPro 155

Qy  562 AGCGTAGACACAGAGGACAGAGGACAGCGGCTATGGAGGAGAGGCCTGAGGAGAGGCCCGAGCAGGAT 621
            |||  |||||  :::::::|||        ||||||:||
Db  156 SerGluAspGluGlnGlyGlnGluGlyThrSerAlaGluSerGlu------ProGluGln--- 172

Qy  622 GGAGTGCAGGTGGCTGTGGTCAGGATCAAAGCCCCCTTGCCCTCCCAGGTAAACAGATTT 681
                       ||:::    :::::|||||
Db  173 -------------LysLysValLysAlaArgArgProValPro--------ArgArgThr 185

Qy  682 ACAGAGAAACTCAACTGCAAAGCCCAACAGAAAATATAGCCCCAGTGGGCAACTTGAAAGG 741
                               |||            |||            |||
```

```
Db  186 ValProLysProArgValLysLysGlnProLys-------ThrGlnArgGlyLysArg 202

Qy  742 AGATGCAGCAGTGGCTGATGAA--                               765
            ::: |||    :::||||||
Db  203 LysLysGlnAspSerSerAspGluAlaProLysAspAspGlyAlaProLysArgGlnThr 222

Qy  766 CACATACAATCCCAGAAGTCAATCCTTTCAGTGAAGAGTTTGATTACGAGGTGGCCATG 825
            |||:|||||||||||||||||||::: |||  :|||||||:||:||
Db  223 ArgArgArgGluAlaAlaLysAsnValSerTyrLysGluAspAspPheGluThrAspSer 242

Qy  826 TCCACCCGCCTACACAAAGGAGGATGAGGGCTATGGCGGCCCCAAGAAGAGGACCAAAACT 885
            |||               :::     :::::|||
Db  243 AspAspLeuIleLeuMetThrGlyGluGlyValAspGlyGlnAspGlnAsnSerGluThr 262

Qy  886 GCTGAAAGG 894
            :::
Db  263 IleGluLys 265
```

```
RESULT 2
GLT4_WHEAT
ID   GLT4_WHEAT     STANDARD;       PRT;     838 AA.
AC   P08489;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutenin, high molecular weight subunit PW212 precursor.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Yamhill;
RX   MEDLINE=86093674; PubMed=3001648;
RA   Sugiyama T., Rafalski A., Peterson D., Soll D.G.;
RT   "A wheat HMW glutenin subunit gene reveals a highly repeated
RT   structure.";
RL   Nucleic Acids Res. 13:8729-8737(1985).
CC   -!- FUNCTION: Glutenins are the high molecular weight seed storage
CC       protein of wheat endosperm. Thought to be responsible for the
CC       visco-elastic property of wheat dough.
CC   -!- SUBUNIT: Disulfide-bridge linked aggregates.
CC   -!- MISCELLANEOUS: Glutenins are coded by several genes on each of the
CC       group 1 chromosomes of wheat.
CC   -!- MISCELLANEOUS: The mature protein is characterized by a large
CC       number of well preserved repeats of the two motifs: GQQPGQ and
CC       GQQPGQGQQGYYPTS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X03346; CAA27052.1; -.
DR   PIR; A24107; EEWTHW.
DR   InterPro; IPR003612; AAI.
DR   InterPro; IPR001419; Glutenin.
DR   Pfam; PF03157; Glutenin_hmw; 1.
DR   PRINTS; PR00210; GLUTENIN.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL        1     21
FT   CHAIN        22    838       GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT                                PW212.
FT   DOMAIN      130    799       REPEATS.
SQ   SEQUENCE   838 AA;  89174 MW;  71D715B7BDF0722D CRC64;

Alignment Scores:
Pred. No.:                0.492          Length:      838
Score:                  123.00          Matches:      75
```

```
Percent Similarity:    34.90%        Conservative:  29
Best Local Similarity: 25.17%        Mismatches:    98
Query Match:            5.98%        Indels:        96
DB:                         1        Gaps:          13

US-10-644-659A-1 (1-1146) x GLT4_WHEAT (1-838)

Qy      9 GGGCGAAAAGGAAAGCAGAAGAGGGAGGCCCAAGAGCGCCCTTCCGGAAGATACGCACAGC 68
            |||::||||||||||:::::  |||   |||
Db    464 GlyGlnLysGlyGlnGlnProGlyGlnGlyGlnProGlyGlnGlyGlnGlnGlnGlnGln 483

Qy     69 CACCCTGGTCATCAGCTTGGCCCGGAGGTTGGCAGCAGTGGGCAATGAGAACAGCATCAG 128
            |||||                                  |||  |||
Db    484 GlnProGly------------------------------GlnGlyGln 489

Qy    129 GCAGGCCCAGGAGCCT-------------ACAGGCTGGCTGCCGGGAGGGACCCAGGACTC 176
            |||  ||||||                  :|  |||||   |||
Db    490 -GlnGlyGlnGlnProGlyGlnGlyGlnProGlyTyrTyrProThrSerProGlnGlnSe 509

Qy    177 ACCTCAAGCTCCTAAACCA---------------------------ATCAC 200
            |||    :::|||
Db    509 rGlyGlyGlyGlnGlnProGlyGlnTrpGlnGlnProGlyGlnGlyGlnProGlyTyrTy 529

Qy    201 ACCCCTACTTCACACCAGAAAGCTCAGAAGTCGCCCAAAGTGCCACCCCGGCCTGCCAGA 260
            |||||||||||||||||
Db    529 rProThrSerProLeuGlnProGlyGlnGlyGlnProGlyTyrAspProThrSerProGl 549

Qy    261 AGGACATGAGAGATGGACAAAGCTCAGAGAAAAGCCCTGAGGTTTCTCACATCAAAAAGAA 320
            :                              |||  |||
Db    549 nGlnProGlyGlyGlnGlyGln---------Pro-GlyGlnLeuGlnGlnProAla 564

Qy    321 AGAGGTGTCCAA--------------AACGGTGTCAGCAAGACTTACGAGAG 359
            ::|||                     |||||||
Db    565 GlnGlyGlnGlyGlnGlnLeuAlaGlnGlyGlnGlnGlnProAlaGlnVal 584

Qy    360 AGGAGGGGACGTGAGCCACCTCAGCCACAGGTA-----------CCGAGAG 398
            :::::||||||||||||||||                      :::::
Db    585 GlnGlnGlyGlnGlnProAlaGlnGlyGlnGlnGlnGlnLeuGlnGlyGlnGlnGlyGln 604

Qy    399 GGATGCTGGTGGTGCTGCTTGAACCTGGGCAGCCAGAGAATGACATTGACAGAATCCTCCACAG 458
            |||                                    ||| |||  |||||
Db    605 Gly--------------------------------GlnGlnProGlyGln 610

Qy    459 CCACGGCTCCCCACGCCGGAGGAGAAAATGTGCCAACCTGGTGTCTCAGCTAACCAAGGG 518
            ::: |||       |||||||
Db    611 GlyGlnGlnProAlaGlnGlyGlnGlnGlyGlnGlnGlnProGly------------ 624

Qy    519 CTGGAGAGTGATGGAGCAGCAGGAGGAGCCCACATGGAGGAGTGACAGCCTAGCACAGAGGA 578
            |||   |||   :|||    :|||:|||:
Db    625 --------GlnGlyGlnGlnGlnGlyGlnGlnProGlyGlnGlyGlnGlnProGlyGlnGly 641

Qy    579 CAGGGGCCATGG------------------AGGAGAGGCTGAGGAGGAGCCCGAGCAGGA 620
            |||                           ::: :
Db    642 GlnProThrTyrTyrProThrSerProGlnGluserGlyGlnGlyGlnGlnProGlyGlnGln 661

Qy    621 TGGAGTGCAGGTGGC-------------TGTGGTCAGGATCAA---------- 650
            |||                         |||||||
Db    662 TrpGlnGlnProGlyGlnTrpGlnGlnProGlyGlnGlyGlnProGlyTyrTyrLeuThr 681

Qy    651 GCGCCCCTTGCCCTC---CCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCA 707
            ::: ||||                        |||||||||||:||
Db    682 SerProLeuGlnLeuGlyGlnGlyGlnGlyGlnGlyTyrProThrSerLeuGlnGlnPro 701

Qy    708 ACAGAAAATATAGCCCAGTGGCAACTTGAAAGGGAGATGGCACAG 753
            :::                    |||:::||||||
Db    702 GlyGln----------GlyGln-GlnProGlyGlnTrpGlnGln 712

RESULT 3
NED4_RAT
ID   NED4_RAT      STANDARD;      PRT;   887 AA.
AC   Q62940;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Ubiquitin-protein ligase Nedd-4 (EC 6.3.2.-).
GN   NEDD4.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND INTERACTION WITH SCNN1A; SCNN1B AND SCNN1G.
RC   TISSUE=Lung;
RX   MEDLINE=96221297; PubMed=8665844;
RA   Staub O., Dho S., Henry P., Correa J., Ishikawa T., McGlade J.,
RA   Rotin D.;
RT   "WW domains of Nedd4 bind to the proline-rich PY motifs in the
RT   epithelial Na+ channel deleted in Liddle's syndrome.";
RL   EMBO J. 15:2371-2380(1996).
RN   [2]
RP   STRUCTURE BY NMR OF 452-499 IN COMPLEX WITH SCNN1B, AND INTERACTION
RP   WITH SCNN1A, SCNN1B AND SCNN1G.
RX   MEDLINE=21223577; PubMed=11323714;
RA   Kanelis V., Rotin D., Forman-Kay J.D.;
RT   "Solution structure of a Nedd4 WW domain-ENaC peptide complex.";
RL   Nat. Struct. Biol. 8:407-412(2001).
CC   -!- FUNCTION: E3 ubiquitin-protein ligase which accepts ubiquitin from
CC       an E2 ubiquitin-conjugating enzyme in the form of a thioester and
CC       then directly transfers the ubiquitin to targeted substrates (By
CC       similarity).
CC   -!- PATHWAY: Ubiquitin conjugation; third step.
CC   -!- SUBUNIT: Interacts with UBE2D2 (By similarity). Binds SCNN1A,
CC       SCNN1B and SCNN1G.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Expression is highest in lung,
CC       kidney and brain.
CC   -!- MISCELLANEOUS: A cysteine residue is required for ubiquitin-
CC       thioester formation.
CC   -!- SIMILARITY: Contains 1 C2 domain.
CC   -!- SIMILARITY: Contains 3 WW domains.
CC   -!- SIMILARITY: Contains 1 HECT-type E3 ubiquitin-protein ligase
CC       domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U50842; AAB48949.1; -.
DR   PIR; S70642; S70642.
DR   PDB; 1I5H; 02-MAY-01.
DR   InterPro; IPR000008; C2.
DR   InterPro; IPR008973; C2_CaLB.
DR   InterPro; IPR000569; HECT_domain.
DR   InterPro; IPR001202; WW_RSP5_WWP.
DR   Pfam; PF00168; C2; 1.
DR   Pfam; PF00632; HECT; 1.
DR   Pfam; PF00397; WW; 3.
DR   PRINTS; PR00360; C2DOMAIN.
DR   SMART; SM00239; C2; 1.
DR   SMART; SM00119; HECTc; 1.
DR   SMART; SM00456; WW; 3.
DR   PROSITE; PS00499; C2_DOMAIN_1; 1.
DR   PROSITE; PS50004; C2_DOMAIN_2; 1.
DR   PROSITE; PS50237; HECT; 1.
DR   PROSITE; PS01159; WW_DOMAIN_1; 3.
DR   PROSITE; PS00020; WW_DOMAIN_2; 3.
KW   Ubl conjugation pathway; Ligase; Repeat; 3D-structure.
FT   DOMAIN       62    163       C2 DOMAIN.
FT   DOMAIN      246    279       WW 1.
FT   DOMAIN      402    435       WW 2.
FT   DOMAIN      459    492       WW 3.
FT   DOMAIN      551    887       HECT.
```

```
FT   BINDING                     UBIQUITIN (BY SIMILARITY).
SQ   SEQUENCE   854    854 AA;  102394 MW;  D74B1097688CD9A1 CRC64;

Alignment Scores:
Pred. No.:
Score:                  0.579
                      122.00
Percent Similarity:    35.85%
Best Local Similarity: 20.28%
Query Match:            5.93%
DB:                           854 AA;

US-10-644-659A-1 (1-1146) x NED4_RAT (1-887)

Qy    34  CCAGCCAAGAGGCGCCCTCCGGAAGATACGCACAGCCACCCTGGTCATCAGCTGGCCCGA  93
               ||| ||||            :::||||| |||||::||||
Db   107  ProMetSerGlyValLeuThrSerValGlnThrLysThrIleLysLysSerLeuAsnPro

Qy    94  GGTTGGCAGCAG----------------------------------------------  105
               ||| :::
Db   107  ProMetSerGlyValLeuThrSerValGlnThrLysThrIleLysLysSerLeuAsnPro  126

Qy   127  LysTrpAsnGluGluIleLeuPheArgValLeuProGlnHisArgIleLeuPheGlu   146

Qy   106  -----------TGGGGAATGAGAACAGCATCAGGCAGCCCAGGAGCCTACA   147
                          |||      ||| |||:  ::: |||
Db   147  ValPheAspGluAsnArgLeuThrArgAspAspPheLeuGlyValAspValProLeu   166

Qy   148  GGCTGGCTGCCGGGAGGGGGACCCAGGACTCACCTCAAGCTCCTAAACCAATCACA-----  201
               ||||||            :::||||||| :::||| |||
Db   167  TyrProLeuPro---------ThrGluAsnProArgMetGluArgProTyrThrPheLys  183

Qy   202  -------------CCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCCAAAGTCGCCACCC  249
                            ||| ||||:              ::: || :||||
Db   184  AspPheValLeuHisProArgSerHis----------LysSerArgValLysGlyTyrLeu  200

Qy   250  CGC----------CTGCCAGAAGGACATGGAGATGGACAAAGCTCAGAGAAAGCC  294
               |||           |||||||:::           :::::::::::||::::|||
Db   201  ArgLeuLysMetThrTyrLeuProLysAsnGlySerAspAspAlaAspGlnAla  220

Qy   295  CCTGAGGTT---------------------------------TCTCACATCAAA  315
               |||:::                                   ::|||:::||:
Db   221  GluGluLeuGluProGlyTrpValValLeuAspGlnProAspAlaAlaThrHisLeuGln  240

Qy   316  AAGAAA----------------------------GAGGTGTCCAAAACGTGGTCAGC  345
               ||||||                            ::: |||:::
Db   241  HisProProGluProSerProLeuProProGlyTrpGluGlnAspValLeuGlnAspValLeuGly  260

Qy   346  AAGACTTACGAGAGAGGAGGGGGACGTGAGCCACCAC----------AGGTACGAGAGG  399
               ::::||||||                           :::::::::::|||:
Db   261  ArgThrTyrTyr----------ValAsnHisGluSerArgThrGlnThrArgProLysArg  276

Qy   400  GATGCTGGTGTGCTTGAACCTGGGCAGCCAGAGAATGACATT-----------  441
               :::                                      ||||:||:||:::
Db   277  ProSer-------------------ProGluAspThrAspGluAsnGly  289

Qy   442  GACAGAAATCCTCCACACGCACCGGCTCCCCCAACGCGGAGGAGAAAAATGTGCCAACCTGGTG  501
               ::||||::|||           :::||| ||   ::: ::
Db   290  AspIleGlnLeuGlnHisAlaHisGlyAlaPheThrArgArgGlnIleSerGluAspVal  309

Qy   502  TCT----------GAGCTAACCAAGGGCTCGAGAGTGATGGAGCAGGAGGAGCAGAGCCC  546
               |||           |||:  || |||||   ::::: :||
Db   310  AspGlyProAspAsnHisGluSerProGluAsnTrpGluIleValAlaArgGluAsnAsn  329

Qy   547  ACATGGAGGAGTGACAGCGTAGACACAGAGGACACGCGGCTATGGAGGAGGAGGTGAGGAG  606
               ||| |||||:::           :::|||:::||
Db   330  Thrile-------------------------TyrSerGlyGlnAlaValGln  338

Qy   607  AGGCCCGAGCAGGAGGATGGAGTGCAGGTGGCGTGTGGTCAGGATCAAGCGCCCCTTGCCCTCC  666
               ||||| ::::
Db   339  SerProProSerGlyHisProAspValGln---ValArgLeuGluAlaGluLeuLeuAspThr  357

Qy   667  CAGGTAAACAGATTACAGAGAAACTCAACTGCAAAGCCC----------AACAGAAATATAGC  720
               ::::: ::                                         :::|||||:::
Db   358  ArgLeuThrMetTyrArgGlyAsnPro-AlaThrSerGlnProValThrSerAsnHisSe   377
```

```
Qy   721  CCAGTGGGCAACTTGAAAGGGAGAATGGCAGCAGTGGCTGATGAACACATACAATCCCAG  780
               :::||                                    :::
Db   377  rSer------------ArgGlyGlySerSerGlnThrCysIlePheGluGlnProTh  393

Qy   781  AAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTCGCCATGTCCACCC--------  832
               |||   |||||
Db   393  rLeuProValLeu------LeuProThrSerSerGlyLeuProProGlyTrpGl  409

Qy   833  -------------------------GCCTACACAAAGAGATGAGGGCTATGGCCGCCCCC  867
                                         :::||||                 :::
Db   409  uGluLysGlnAspAspArgGlyArgSerTyrTyr-ValAspHisAsnSerHisThrThrh  429

Qy   868  AAAGAAGGAACCAAAACTGCTGAAAGGGCA--AGGGTGCTGAGGAGCACATCTACAGG  924
               :::|||                         :::        |||||||||:::|||
Db   429  rTrpSerLysProThrMetGlnAspAspArgSerLysIle-ProAlaHisLeuArgG  449

Qy   925  GA----------------------------------------------------------  926
               ||
Db   449  lyLysThrProValAspSerAsnAspLeuGlyProLeuProProGlyTrpGluArgT  469

Qy   927  --AATGATGGACATGTCCTTCATTATCTGCACAATGGCTCGCCACGACGAGATGGC  981
                 :::||||                 ::: :::||::||:
Db   469  hrHisThrAspGlyArgValPheValPhePheIleAsnHisAsn---IleLysLysThrGlnTrpG  488

Qy   982  AAGATCCA  989
               :::|||||
Db   488  luAspPro  490
```

```
RESULT 4
MAPA_HUMAN
ID   MAPA_HUMAN      STANDARD;      PRT;   2805 AA.
AC   P78559; O95643; Q12973; Q15882; Q9UUT4;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1A (MAP 1A) (Proliferation-related
DE   protein p80) [Contains: MAP1 light chain LC2].
GN   MAP1A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97001161; PubMed=8812494;
RA   Fink J.K., Jones S.M., Esposito C., Wilkowski J.;
RT   "Human microtubule-associated protein 1a (MAP1A) gene: genomic
RT   organization, cDNA sequence, and developmental- and tissue-specific
RT   expression.";
RL   Genomics 35;577-585(1996).
RN   [2]
RP   SEQUENCE OF 1-1825 FROM N.A.
RC   TISSUE=ovarian carcinoma;
RA   Chen Z.C., Fadiel A., Naftolin F.;
RT   "Identification of a novel protein (p80) in ovarian carcinoma cells.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 78-1687 FROM N.A.
RC   TISSUE=Brain;
RA   Ohtani K., Rutherford T., Sakamoto H., Naftolin F.;
RT   "Microtubule associated protein 1A (MAP1A) in human brain - DNA
RT   sequence and physiological role.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 134-419 FROM N.A.
RC   TISSUE=fetal muscle;
RA   Chiannilkulchai N., Pasturaud P., Richard I., Auffray C.,
RA   Beckmann J.S.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1607-1883 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95356255; PubMed=7629894;
```

RA   Fukuyama R., Rapoport S.I.;
RT   "Brain-specific expression of human microtubule-associated protein 1A
RT   (MAP1A) gene and its assignment to human chromosome 15.";
RL   J. Neurosci. Res. 40:820-825(1995).
CC   -!- FUNCTION: Structural protein involved in the filamentous cross-
CC   bridging between microtubules and other skeletal elements.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC   with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: Brain.
CC   -!- DOMAIN: THe basic region containing the repeats may be responsible
CC   for the binding of MAP1A to microtubules.
CC   -!- PTM: Various serine residues may be phosphorylated by cAMP kinase.
CC   -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC   FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC   BOTH MAP1A AND MAP1B.
CC   -!- SIMILARITY: TO MAP1B.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U38291; AAB41132.1; -.
DR   EMBL; U38292; AAB41133.1; -.
DR   EMBL; AF200415; AAF08305.2; -.
DR   EMBL; U80458; AAD00355.1; -.
DR   EMBL; Z47038; CAA87104.1; -.
DR   EMBL; U14577; AAA81362.1; -.
DR   PIR; I38857; I38857.
DR   Genew; HGNC:6835; MAP1A.
DR   MIM; 600178; -.
DR   GO; GO:0005875; C:microtubule associated complex; TAS.
KW   Microtubule; Repeat; Phosphorylation.
FT   CHAIN   72490   2805   MAP1 LIGHT CHAIN LC2.
FT   DOMAIN   309   496   LYS-RICH (BASIC).
FT   DOMAIN   415   541   9 X 3 AA REPEATS OF K-K-[DE].
FT   REPEAT   415   417   1.
FT   REPEAT   420   422   2.
FT   REPEAT   427   429   3.
FT   REPEAT   431   433   4.
FT   REPEAT   436   438   5.
FT   REPEAT   440   442   6.
FT   REPEAT   444   446   7.
FT   REPEAT   449   451   8.
FT   REPEAT   539   541   9.
FT   CONFLICT   72   72   F -> L (IN REF. 1).
FT   CONFLICT   134   135   VV -> IP (IN REF. 4).
FT   CONFLICT   249   249   A -> G (IN REF. 4).
FT   CONFLICT   263   263   V -> A (IN REF. 4).
FT   CONFLICT   296   296   Q -> H (IN REF. 3).
FT   CONFLICT   311   311   S -> G (IN REF. 4).
FT   CONFLICT   324   324   K -> Q (IN REF. 3).
FT   CONFLICT   335   336   AK -> ST (IN REF. 1).
FT   CONFLICT   353   353   A -> S (IN REF. 1).
FT   CONFLICT   357   357   A -> S (IN REF. 1).
FT   CONFLICT   364   364   K -> Q (IN REF. 1).
FT   CONFLICT   414   419   EKKDKE -> KKKRNS (IN REF. 4).
FT   CONFLICT   424   424   K -> P (IN REF. 1).
FT   CONFLICT   426   426   E -> D (IN REF. 3).
FT   CONFLICT   431   431   K -> Q (IN REF. 3).
FT   CONFLICT   439   439   E -> D (IN REF. 3).
FT   CONFLICT   444   444   K -> R (IN REF. 3).
FT   CONFLICT   452   453   TK -> SS (IN REF. 3).
FT   CONFLICT   457   457   K -> R (IN REF. 3).
FT   CONFLICT   682   682   Q -> P (IN REF. 1).
FT   CONFLICT   1025   1025   Q -> K (IN REF. 1).
FT   CONFLICT   1303   1313   KYLPGAITSPD -> EVLTWGDHQALN (IN REF. 3).
FT   CONFLICT   1335   1341   MISSING (IN REF. 3).
FT   CONFLICT   1368   1368   Q -> T (IN REF. 3).
FT   CONFLICT   1470   1470   A -> T (IN REF. 3).
FT   CONFLICT   1650   1650   W -> C (IN REF. 1).
FT   CONFLICT   1690   1690   A -> S (IN REF. 1).
FT   CONFLICT   1714   1714   G -> V (IN REF. 1).
FT   CONFLICT   1869   1869   E -> A (IN REF. 5).
FT   CONFLICT   1879   1883   GTPEY -> AHSRV (IN REF. 5).
SQ   SEQUENCE   2805 AA; 306392 MW; F3ED0A3165993B2E CRC64;

Alignment Scores:
Pred. No.:   Length: 2805
Score: 0.761   Matches: 77
Percent Similarity: 121.00   Conservative: 31
Best Local Similarity: 33.86%   Mismatches: 133
Query Match: 24.14%   Indels: 78
DB: 5.88%   Gaps: 12
                                 1

US-10-644-659a-1 (1-1146) x MAP1A_HUMAN (1-2805)

Qy   48 CCTCCGGAAGATACGCAC--AGC   68
            |||||||||||:::   :::
Db   1781 ProProGluGluAspLysLeuThrArgSerProPheGlueIleIleSerProProAla   1800

Qy   69 CACCCTGGTCATCAGCTTGGCCCCAGGTTG---GCAGCAGTGGCGAATGAGAACAGCAT   125
            ||| |||:::|||::: |||::: |||||||
Db   1801 SerProProGluMetValGlyLnArgValProSerAlaProGlyGln-GluSerProI   1820

Qy   126 C----------------------------AGGCAGGCCCAGGAGCCTACA---------GGCTGGCTG-   156
            | |||||||
Db   1820 eProAspProLysLeuMetProHisMetLysAsnGlnProThrProSerTrpLeuAl   1840

Qy   157 ----------CCGGGAGGACCCAGGACTCACCTCAAGCTCCTAAACCAATCACACCCC   206
                      ||||||||||||||||||||||||||||||||||||::::|||||
Db   1840 aAspIleProProTrpValProLysAspArgProLeuProProAlaProLeuSerPro-   1859

Qy   207 TACTTCACACAGAAAGCTCAGAGTGCCCCAAAGTCGCACCCCGCCTGCCAGAAGGACA   266
            ||||| |||||||||||||||||
Db   1860 ----------AlaProGlyProThrProAlaProGluSerHi   1871

Qy   267 TGGAGATGACAGAATCCTCCACAGCCAGGCTCCCAACGCGGAGGAGGAGAAAATGTGC   326
            ||||| ||||||||||||||||||||||||||||||||||
Db   1871 sThrProAlaProPheSerTrpGlyThrProGluTyrAsp-----------------   1884

Qy   327 GTCCAAAACGGTGGTCAGCAAGACTTACGAG------------------AGAGGAGGGACGT   371
            ::::|||||:::
Db   1885 ------SerValValAlaAlaValGlnAlaValGluAlaAlaGluLeuGluGlyProTy   1902

Qy   372 GAGGCACCTCAGCCACAGGTACGAGAGGGGATGCTGGTGCTTGAACCTGGGCAGCCAGA   431
            ||||||||||||||||| |||::: |||
Db   1902 rSerProLeuGly-LysAspTyrArgLysVal***GlyGluArgGluGluGluGlyArgAl   1922

Qy   432 GAATGACATTGACAGAATCCTCCACAGCCAGGCTCCCAACGCGGAGGAGGAGAAAATGTGC   491
            |||||||||||||||||||||
Db   1922 aGluAlaProAspLysSerSerHis***SerLysValProGluAlaArgLysSerHisAl   1942

Qy   492 CAACCTGGTCTCTGAGCTAACCAAGGGCTGGAGAGTGATGAGCAGCAGGAGCCCACATG   551
            :::||||||||||
Db   1942 aThrThrGluProGluGlnThrGlu---------ProGluGlnArgGluProThrPr   1958

Qy   552 GAGGAGTGACAGCGTAGACACAGAGGACACAGCGCTATGGAGGAGGGCTGAGGAGAGGCC   611
            |||
Db   1958 oTyrProAspGluArgSerPheGln-TyrAlaAspIleTyrGluGlnMet----------   1974

Qy   612 CGAGCAGGATGGAGTGCAGGTGGCTGGTCAGGATCAAGCGCCCCTTGCCCCCCAGGT   671
            ::: :::|||:||
Db   1975 -----------MetLeuThrGlyLeuGlyProAlaCysProThrArgG   1987

Qy   672 AAACAGATTTACAGAGAAACTCAACTGCAAAGCCAACAGAAAATAGCCCAGTGGGCAA   731
            |||:::||||||||||| ::: |||||||
Db   1987 luProProLeuGlyAlaAlaGlyAspThrProProCysLeuSerThrLysGluAlaAlaA   2007

Qy   732 CTTGAAAGGAGAGATGGCAGCAGTGGGCTGATGAACATGACACATACAATCCAGAAGCTCAATCC   791
            ::: |||||
Db   2007 laGlyArgAsnThrSerAlaGlu-----------------------------LysGluLeuSerProI   2021

```
Qy    792  TTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAGGAGATGA 851
           ::||| ||||||||| ||||||    ::: |||                      |||
Db   2021  leSerProLysSerLeuGlnSerAspThrPro------------------ThrPheSerTyrA 2036

Qy    852  GGGCTATGGCCGCCCAAAGAAGAACCAAAACTGCTGAAAAGGGCCAAGCGTG 904
           ||||:::|||                  ||||||||:::
Db   2036  laAlaLeuAlaGlyProThrValProProArg***GluProGlyProSerMet 2053
```

RESULT 5

```
GLT5_WHEAT
ID   GLT5_WHEAT      STANDARD;      PRT;      839 AA.
AC   P10388;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glutenin, high molecular weight subunit DX5 precursor.
GN   GLU-1-1D OR GLU-D1-1B.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Cheyenne;
RX   MEDLINE=89098419; PubMed=2563152;
RA   Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
RA   Malpica-Romero J.M.;
RT   "Nucleotide sequences of the two high-molecular-weight glutenin genes
RT   from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
RT   Cheyenne."
RL   Nucleic Acids Res. 17:461-462(1989).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RC   STRAIN=cv. Cheyenne;
RA   Anderson O.D.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Glutenins are the high molecular weight seed storage
CC       protein of wheat endosperm. Thought to be responsible for the
CC       visco-elastic property of wheat dough.
CC   -!- SUBUNIT: Disulfide-bridge linked aggregates.
CC   -!- MISCELLANEOUS: Glutenins are coded by several genes on each of the
CC       group 1 chromosomes of wheat.
CC   -!- MISCELLANEOUS: The mature protein is characterized by a large
CC       number of well preserved repeats of the two motifs: GQPGQ and
CC       GQPGQGQGYPTS.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; X12928; CAA31395.1; -.
DR   InterPro; IPR001419; Glutenin.
DR   Pfam; PF03157; Glutenin_hmw; 1.
DR   PRINTS; PR00210; GLUTENIN.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL      1    21
FT   CHAIN      22   839       GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT
FT                             DX5.
FT   DOMAIN    131   801       REPEATS.
SQ   SEQUENCE   839 AA;   89359 MW;   0F1431106D552643 CRC64;
```

Alignment Scores:

```
Pred. No.:                         Length:          839
Score:             0.787           Matches:          92
Percent Similarity: 36.39%         Conservative:     39
Best Local Similarity: 25.56%      Mismatches:      125
Query Match:       120.00          Indels:          104
                                   Gaps:             19
```

```
US-10-644-659A-1 (1-1146) x GLT5_WHEAT (1-839)

Qy      9  GGGCGAAAGGAAAGCGGGGAGGGCGAGGGCCAAGCAGCCGCCTCCGGAGATACCCACAGC 68
           |||:::||||||||||:::::||||| ||| ||||| ||:::
Db      1  GlyGlnLysGlyGlnInProGlyGlnGlyGlnInProGlyGlnGlyGlnInGlyGlyGln 478

Qy     69  CACCCTGGTCATCAGCTTGGCCCCAGAGTTGGCAGTGGCCAGTGGCGAATGAGAACAGCATCAG 128
           ||| ::||| |||
Db     69  GlnGlyGln 484

Qy    129  GCAGGGCCAGAGCCT-----------ACAGGGCTGGCTGCCGGGAGGGACCCAGGACTC 176
           ||| ||::: ||                    :||| ||| |||
Db    485  -GlnGlyGlnInProGlyGlnGlyGlnInProGlyTyrTyrProThrSerProGlnGlnSe 504

Qy    177  ACCTCAAGCTCCTAACCA--------------------------------ATCAC 200
           :|||
Db    504  rGlyGlnInGlyGlnInProGlyGlnInTrpGlnInProGlyGlnGlyGlnProGlyTyrTy 524

Qy    201  ACCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAGTCGCCACCCGCCTGCCAGA 260
           ||| :::            |||||
Db    524  rProThrSerProLeuGlnProGlyGlnGlyGlnInProGlyTyrAspProThrSerProGl 544

Qy    261  AGGACATGGAGATGGACAAAGCTCAGAGAAAGCCCTGAGGTTTCTCACATCAAAAAGAA 320
           :                     |||||||
Db    544  nGlnProGlyGlnGlyGlnGln-----------Pro--GlyGlnLeuGlnInProAla 559

Qy    321  AGAGGTGTCCAAAACGGTGCTCAGCAAGACTTAGCAGAGGAGGGGACGTGAGCCACCT 380
           :::||||||                    |||||||:::|||
Db    560  GlnGlyGlnGln-------GlyGlnInLeuAlaGlnGlyGlnGlyGlnInGlyGlnInPro--- 576

Qy    381  CAGCCACAGGTACGAGAGGGGATGCTGGTGTTGCTTGAACCTGGCAGCCAGAGAATGACAT 440
           |||||||||:::::|||
Db    577  ---AlaGlnValGlnGlyGly- 582

Qy    441  TGACAGAATCCTCCACAG-----------CCACGGCTCCCCAACGCGGAGGAGGAGAAAATG 488
           :::|||
Db    583  ---GlnGlnProAlaGlnGlnGlyGlnGlnLeuGlyGlnGlyGlnGlyGlnGln 601

Qy    489  TGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGAGGAGCCCAC 548
           |||||||                 ||| |||
Db    602  ---GlnProGly-------GlnGly-----GlnGlnGlyGlnGlnProAlaGln 614

Qy    549  ATGGAGAGTGACAGCGTAGACACAGGAGACAGCGGCCATGGAGGAGGGCTGAGGAGAG 608
           :::|||||| :::||||                 ||| |||||||||
Db    615  GlyGlnGlyGlnInGlyGlnInProGlyGlnGlyGlnInHisGlnGlnProGlyGlnGlnGln 634

Qy    609  GCCCGGAGCAGGAGTGGAGTGCAGGTGCGTGGTCAGGATCAAGCGCCCCTTGCCCTCCCA 668
           |||||||||                ||||||||
Db    635  ----------------GlnGlyGlnInProGlyGlnInGlnInPro----GlyGln 647

Qy    669  GGTAAACAGATTTACAGAGAAACTCAACTGCAAGCCCAACAGAAATATAGCCAGTGGG 728
           |||::|             |||||||
Db    648  GlyGlnProTrpTyrTyrProThr-------SerPro-GlnGlu--SerGlyGlnGl 663

Qy    729  CAACTTGAAAGGAGAGATGGCAGCAGTGGGCTGATGAACACATACAATCCAGAGCTCAA 788
           |                      |||:::||||||||
Db    663  yGlnGlyGlnInProGlyGlnInTrpGlnGln- 671

Qy    789  TCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAGGAGA 848
           :::                       |||
Db    672  -ProGlyGlnInGlyGlnProGlyTyrTyrLeuThrPheSerValAlaLeuAlaArgThrGlyGl 691

Qy    849  TGAGGGCTAT------------------------GGCCGCCCAAAGAAGAACCAAAACTGCTGAAAG 893
           ||:::                            :::|||
Db    691  nGlnGlyTyrTyrProThrSerLeuGlnGly------------------GlnGl 707

Qy    894  GGGCAAGCGTGCTGAGGAGCACATCTCAGGGAAATGATGG----------ACATCTGCTT 944
           ||||||                                           :::|||
Db    707  n-ProGlyGlnInTrpGlnInSerGlyGlnGlyGlnInHisTrpTyrTyrProThrSerProL 727

Qy    945  CATTATCTGCACAATGCTCCGCCACAGACCAGATGCAAGATCCAGGT 992
           |||||| |||   :::||| |||       ||||||
Db    727                                                     GlnGl 707
```

```
Db    727 ysLeuSerGlyGln-GlyGlnArgProGlyGlnTrpLeuGlnProGly 742

RESULT 6
HAIR_MOUSE
ID   HAIR_MOUSE      STANDARD;      PRT;   1182 AA.
AC   Q61645; Q80Y47;
DT   01-NOV-1997 (Rel. 35, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hairless protein.
GN   HR.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Skin;
RX   MEDLINE=9329587; PubMed=8052649;
RA   Cachon-Gonzalez M.B., Fenner S., Coffin J.M., Moran C., Best S.,
RA   Stoye J.P.;
RT   "Structure and expression of the hairless gene of mice.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:7717-7721(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Retina;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May act as a transcription factor that could act on to
CC       regulate one of the phases of hair growth.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed predominantly in brain, hair
CC       follicles and interfollicular epidermis. No expression in
CC       dermis.
CC   -!- DISEASE: HR MUTATION PRODUCES A NUMBER OF PLEIOTROPIC EFFECTS
CC       INCLUDING STRUCTURAL ABNORMALITIES OF EPITHELIAL CELLS IN THE HAIR
CC       FOLLICLES, HAIR LOSS TOWARDS THE END OF THE FIRST HAIR GROWTH
CC       CYCLE, AND THE FAILURE OF SUBSEQUENT HAIR GROWTH CYCLES. OLDER
CC       MICE CARRYING AN HR MUTATION HAVE BEEN REPORTED TO POSSESS ALTERED
CC       RATIOS OF T-CELL-DEPENDENT B-CELL RESPONSES. MICE HOMOZYGOUS FOR
CC       HR MUTATION ARE UNIQUELY SENSITIVE TO UV AND CHEMICALLY INDUCED
CC       SKIN TUMORS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z32675; CAA83587.1; -.
DR   EMBL; BC049182; AAH49182.1; -.
DR   PIR; I48378; I48378.
```

```
DR   MGD; MGI:96223; hr.
DR   InterPro; IPR003347; TF_JmjC.
DR   Pfam; PF02373; jmjC; 1.
DR   SMART; SM00558; JmjC; 1.
KW   Zinc-finger; DNA-binding; Nuclear protein; Transcription regulation;
KW   Metal-binding.
FT   ZN_FING       595    620       C6-TYPE.
FT   DOMAIN        535    540       POLY-GLY.
FT   CONFLICT      401    401       P -> S (IN REF. 1).
SQ   SEQUENCE   1182 AA;  127192 MW;  3AFABE96C6EB3241 CRC64;

Alignment Scores:
Pred. No.:                               Length:            1182
Score:                 120.00            Matches:             98
Percent Similarity:     35.94%           Conservative:        40
Best Local Similarity:  25.52%           Mismatches:         114
Query Match:             5.83%           Indels:             132
DB:                                      Gaps:                25

US-10-644-659A-1 (1-1146) x HAIR_MOUSE (1-1182)

Qy    23 GCGGGGAGGGCCCAG---------------------------------CCAAGA   43
          ||||||||||||||||                                ||||||
Db   255 AlaGlyArgGlnGlnAsnLeuCysProValPheLeuGlyTyrProAspThrValProArg  274

Qy    44 GCGCCCTCCGGAAGATACGGCCACAGCCACCCTGGT-------CATCAGGTTGGCCCGAGGTT   97
          ||||||                       :::      |||  |||||||
Db   275 AlaProTrpPro------SerCysPro-ProGlyLeuValHisSerLeuGly--------   289

Qy    98 GGCAGCAGTGGGCGAATGAGAACAGCATCAGGCAGGCCCAGGAGCC---------TA  145
          |||||||||||||||||||||||||||||||||||||||||||
Db   290 ----------------------AsnIleTrpAlaGlyProGlySerAsnSerLeuGlyTy  302

Qy   146 CAGGCTGGCTGCCGGGAGGGACCCAGGACTCACCTCAAGCTCCTAAACCAATCACACCCC  205
          |:::||||       ||||          |||  ||  :::||||  |||||
Db   302 rGlnLeuGly-ProProAlaThrProArgCysPro---SerProGlyProProThrProP  321

Qy   206 CT------------ACTTCACACAGAAGCTCAGAGTGCCCAAAGTCGCCACCCGC-  252
          |||            ||||||||||||||||||||||||||||||
Db   321 roGlyGlyCysSerSerHisLeuProAlaArgGluGluGlyAspLeuGlyProCysArgL  341

Qy   253 -CTGCAGAAGGA------------CATGAGAATGGACAAAGCTCAGAGAAA---  291
          ||||||||                |||||||||||||:::||||||||
Db   341 ysCysGlnAspSerProGluGlyGlyGlySerSerGlyProGlyGluSerSerGluGluArgA  361

Qy   292 --------GCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCA  331
                  |||||||||||||||||||||||||||||||:::
Db   331 ------GCCCCTGAGGTTTCTCACATCAAAAAGAAAGAGGTGTCCA  331

Qy   361 snLysAlaAspSerArgAlaCysProProSerHisThrLysLeuUysThrLysTrpTrpL  381
Db   381 AGTTACGAGAGGGGATGCTGAGTGTCGTGTGCTTGAA------------------------   381

Qy   332 AAACGGTGGTCAGCAAGACTTACGAGAGGAGGGAGCGTGAGCCACCTCAGCCAC---  387
          ||||||||||||||||       :::||||       |||
Db   401 roAlaThrGlyLeuArgAlaLeuLysArgAlaAlaGlySer-ProGluValGlnGlyAlaSerA  421

Qy   381 euThrArgHisSerGluGlnPheGluCysProGlyGlyCysProGlyLysSerGlyLysGluSerP  401
          :::||||       ||||
Db   401 ----CCTGGCCGGCCAGGAAGGATGACATTGACAGAATCCTCCACAGCCACGGCT  466

Qy   388 --------AGGTACGAGAGGGATGCTGAGTGTCGTGTGCTTGAA-  417
                  |||:::                  :::
Db   421 rgGlyProAlaProLysArgPro---------------------SerHisProP  432

Qy   467 CCCCAACGCGGAGGAGAGAAAATGTGCCAACCTGGTGTCTGAGCTAACCAAGGGCTGGAGAG  526
          |||       ||||||:::|||  |||
Db   432 heProGlyThrGlyArgGlnGlyAla---------------------ArgAlaTrp----  443

Qy   527 TGATGGAGCAGGAGGAGCCCACATGGAGTGACAGCCGTAGCACAGAGGACAGCGGCT  586
          |||||||||||||           ::::::::::
Db   444 ----------GlnGluThrPro---------------------GluGlyS  453

Qy   587 ATGGAGGAGAGGCTGAGGAGGAGGCCCGAGCAGGAGGATGGAGGAGCTGTGGCTGTGGTCAGGA  646
          |||||||||||:::::|||||  |||          |||  :::
Db   453 erLysAlaGluAlaGluGluGlnGluGlnGlnLeuGlnLeuArgGlyProArgAspGlyArgIleArgL  473
```

```
SQ   SEQUENCE   458 AA;  52768 MW;  10CF968FFDAA0C69 CRC64;

Alignment Scores:
Pred. No.:
Score:                   119.50                 Length:         458
Percent Similarity:       38.36%                Matches:         89
Best Local Similarity:    23.54%                Conservative:    56
Query Match:               5.81%                Mismatches:     161
DB:                        1                    Indels:          72
                                                Gaps:          '19

US-10-644-659A-1 (1-1146) x YNE1_CABEL (1-458)

Qy  115 GAGAACAGCATCAGGCAGGCCCAGGAGCCTACAGGCTGGCTGCGGGAGGAGGACCCAG--  171
            ::   |||||  |||:|:||||:|||||:|:::
Db   68 LysGlnSerValLysGlyAlaLysValValHisGlyPheLysGlnThrPheAlaGluVal   87

Qy  172 GACTCACCTCAAGCT-----CCTAAACCAATCACACCCCTACTTCA-------         213
            ||::|||::          |||::||||::||:||:|:||||||
Db   88 AspAlaProLysValIleLysProGluProLeuSerProAspAsnSerProArgGly     107

Qy  214 ------------CACCAGAACAGCTCAGAGTGCCCCAAG-------------        240
Db  108 LysArgGlnHisAspSerAspAsnSerProProArgProSerArgLysArgAsnAsp     127

Qy  241 -------TGCCACCCGCTGCCAGAGGACATGGACGATGGACAAAGCTCAGAGAAA        291
            ||||||||
Db  128 SerAspAsnSerProProArgProSerArgAsnArgHisAspSerAspLysAspAsnSer   147

Qy  292 GCCCCTGAGGTTTCTCACATCAAAAGAAAGAGAGTGTCAAAACGTGGTCAGCAAGACT    351
            :::  |||
Db  148 ProPro-ArgArgArgHisAspSerAspAsnSerProProSerArgLys-I          167

Qy  352 TACGAGAGAGG------------AGGGGACGTGAGCCACCTCAGCCACAGGTACGAGAGG  399
            ::||||||                   |||  ||||||  ||| |||
Db  167 IeArgGluGluSerProSerAlaArgAsnArgArgSerPro---ProArgThrArg-Arg  185

Qy  400 GATGCTGGTGTGTGCTTGAACCTGGGCAGCCAGAGAAT-----------GACATT      441
            :::                               |||
Db  186 AspArgHisAspSerAspAsnSerProProArgAsnArgSerArgAspSerAspAsn     205

Qy  442 GACAGAATCCTCCACAGCCAGCCGCCTCCCAACGGGAGGAGAAAATGTGCCAACCTGGTG  501
            |||||||||||                    ||||    :::||
Db  206 SerProProArgArgProSerSerProAlaArgArgArgLysAspAspAspLeu---     224

Qy  502 TCTGAGCTAACCAAGGGCTGGAGAGTGATGGAG----------------CAGGAGGAGCC  546
            |||  ||| |||              :::||||||||
Db  225 SerProProArgLysSerArgLysIleGluProLysLysIleLysGluGluPro        244

Qy  547 ACATGGAGGAGTGACACGCTAGAACACAGAGACACGGCTATGGAGGAGAGGCTGGAGGAG  606
                    :::                            :::
Db  245 --------AspSer---AspThrGluThrSerGlyArgThrLeuGluGly-------     257

Qy  607 AGGCCCGGAGCCAGCAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCCTCC 666
            ||| |||  |||                    |||  |||
Db  258 -----LysArgSerGlyLeuGlnSerAla------------ArgAspLeuLysGlu      270

Qy  667 CAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAACAGAGAAATATAGCCCAGTG 726
            ::: |||                    :::   ||
Db  271 Glu------------SerAspLysLeuArgAlaLysAsnSerLysMetPheGluGluMet   286

Qy  727 GGC---AACTTGAAAGGAGGATGGCAGCAGTGGGCTGATGAACACATACAATCCCAGAAG  783
            |||  :::                           |||:::|||||
Db  287 AspThrSerValSerGlyArgPheAlaAspThrValTyrArgGlnLysGlnThrLysLys  306

Qy  784 CTCCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACC---CGCTACAAC 840
            :::                                          |||
Db  307 LysGlyLysAspSerGluGluLysGlnAlaLysLysGluArgGluThrLysLysThrGlu  326

Qy  841 AAAGGAGGATGGAGGCTATGGCCGCCCCCAAAGAAGAAGACCAAAACTGCTGAAAGGGCCAAG 900
            :::                               |||
Db  327 GluLeuLysGluLysTyrLysSerTrpAsnLysGlyValAlaGlnIleIleGluAspArgArg  346

Qy  901 CGTGCTGGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATG  960
```

```
Qy  647 TCAAGCGCCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTGCAAA---  702
            ::   :::                                  :::||||
Db  473 eu----------GlnGluSerArgLeuValAsp---ThrSerCysGlnHisHisH     486

Qy  703 ----------GCCCAACAGAGAAATAGCCCAGTGGCAACT                     733
                      ||||
Db  486 isLeuAlaGlyValThrCysGlnSerValGlnAlaAlaGlyValGlyValL           506

Qy  734 TGAAAGGGAGGATGGAGCAGTGGGCTGATGAACACATACAATCCCAGAAGCTCAATCCTT  793
            |||                                     ||| :::||
Db  506 euThrGly--------------HisSerGlnLysSerArgArgSerProL           518

Qy  794 TCAGTGAA--GAGTTTGATTACGAGCTGGCCATGTCCACCCGCCTACACAAAGGAGATG   850
            :::  |||    :::  ||                            :::
Db  518 euGluGlyLysGlnLeuGluGluGluLysAspSerSerAlaThr---------SerGluG  534

Qy  851 AGGGCTATGGCCGCCCCCAAAGAAGAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCTGAGG 910
            |||                                |||
Db  534 luGlyGlyGlyPro--------GlyProGluAlaSerLeuAsnLysGlyLeuAlaL       551

Qy  911 AGCACATCTACAGGGAAAATGATGACATGTGCTTCATTATCTGCACAATGGCTCGCCACA   970
            ::|||:::                         :::
Db  551 ysHisLeuSerGlyLeuGlyLeuAspArg------LeuCysArgLeuLeuArgLysG    568

Qy  971 GACGAGAT                                                      978
            |||:::
Db  568 luArgGlu                                                      570
```

```
RESULT 7
YNE1_CABEL
ID   YNE1_CABEL     STANDARD;      PRT;   458 AA.
AC   P30640;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Hypothetical protein R08D7.1 in chromosome III.
GN   R08D7.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showrkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368;32-38(1994).
-!- SIMILARITY: SOME, TO YEAST YGL174W.
CC   -----------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; Z12017; CAA78047.1; -.
DR   PIR; S24457; S24457.
DR   WormPep; R08D7.1; CE00289.
KW   Hypothetical protein.
```

Alignment Scores:
Pred. No.: 1.02
Score: 118.50
Percent Similarity: 32.04%
Best Local Similarity: 20.61%
Query Match: 5.76%
DB:

Length: 1067
Matches: 101
Conservative: 56
Mismatches: 136
Indels: 197
Gaps: 25

US-10-644-659A-1 (1-1146) x MICA_HUMAN (1-1067)

```
Qy    31 GGGCCCAGCCAAGAGCGCCTCCGGAAGATACGGACAGCCAGCACCCTGGTCATCAGC---TTG   87
            |||||||                :::           ||||||            :::|||    |||
Db   619 GlyProValSerGlnAlaSerProGlyThrSerProAlaValLeuPheLeuSerLysLeu  638

Qy    88 GCCCGAGGTTGGCAGCAG---TGGGCGAATGAGAACAGC-                        123
            ||||  |||:::                |||||||||||  ||||:
Db   639 GlnArgThrLeuGlnArgSerArgArgAlaLysGluAsnAlaGlyAlaGlyLysLysLys  658

Qy   124 ATCAGG----------CAGGCCCAGGAGCCTACAGGCTGGCTGCGCGGAGGGGACCCAGGAC  174
            ::::||          |||||:::  ||||| |||:::   ||||||
Db   659 LeuArgLeuMetGluAlaGluHisProSerThrGluValPro--------------Pro  674

Qy   175 TCACCTCAAGCTCCTAAACAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGT--  231
            |||:::     |||||||||||||||||||  ||||||||:::|||
Db   675 AspProGluProGlyValProLeuThrProSerGlnHisGlnHisGluAlaGlyAlaGly  694

Qy   231 ---                                                          231
Db   695 AspLeuCysAlaLeuCysGlyGluHisLeuTyrValLeuGluTyrLeuCysValAsnGly  714

Qy   232 ---GCCCAAAGTCGCCACCC  249
            |||||  |||
Db   715 HisPhePheHisArgSerCysPheArgCysHisThrCysGluAlaThrLeuTrpProGly  734

Qy   250 CGCCTGCCAGAAGGACATGGAGATGGACAA-------AGCTCAGAGAAAGCCCCTGAGGTT  303
            :::  ||||||      ||||||||  :::
Db   735 GlyTyrGluGlnHisProGlyAspGlyHisPheTyrCysLeuHisLeuProGlnHisThr  754

Qy   304 TCTCAACATCAAAAAGAAAAAGAAGAGTGTCCAAAACGGTGTCAGACTTACGAGGAGAGGA  363
                         :::||||||
Db   755 AspHis-----------LysLysGluGlySer-----------------AspArgGly  764

Qy   364 GGGGACGTGAGCCACCTCAGCCACCAGGTAGGAGAGGGAT-------------------  402
                :::   |||
Db   765 ProGluSerProGluLeuProThrProSerGluAsnSerMetProProGlyLeuSerThr  784

Qy   403 -----------------------GCTCGT--GTGCTTGAACCTGGGCAGCCAGAGAATGAC  438
                                      ||||||  |||||
Db   785 ProThrAlaSerGlnHisGluGlyAlaGlyProValProSerGlnProThrArgArgArg  804

Qy   439 ATTGACAGAATCCTCCACAGCCACCACGGCTCCCCAAGCGGAGG---------------  480
            |||||      :::|||||||:::
Db   805 GlnIleArgLeu---------------SerSerProGluArgGlnArgLeuSerLeuAsn  820

Qy   481 ----------------------------------AGAAAATGTGCCAACCTG  498
                                                ||| :::|||
Db   821 LeuThrProAspProGluMetGluProProLysProProArgSerCysSerAlaLeu  840

Qy   499 GTGTTCTGAGCTAACCAAG-----------------GGCTGGAGAGTGATGGAGCAGGAGGAGCCC  546
            |||           ||| :::
Db   841 AlaArgHisAlaLeuGluSerPheValGlyTrpGlyLeuProValGlnSerProGln  860

Qy   547 ACATGGAGGAGTGACAGCGTAGACACAGAGACAGGCGCTATGGAGGAGGCTGACGAG  606
                      :::||||:::      ::::::|||
Db   861 AlaLeuValAlaMetGluLysGluGluLysGluSerProPheSerProGluGluGlu  880

Qy   607 AGG--------CCCGAGCAGGATGGAGTGGCAGGTGGCTGTG-----------  639
            |||        :::|||
Db   881 GluAspValProLeuAspSerArgValGluGlnAlaLeuGlnThrPheAlaLysThrSer  900

Qy   640 GTCAGGATCCAAGCAGCCCCTTGCCCTCC  666
                                 :::::::|||
```

RESULT 8
MICA_HUMAN
ID   MICA_HUMAN   STANDARD;   PRT;   1067 AA.
AC   Q8TDZ2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   CasL interacting molecule (NEDD9 interacting protein with calponin
DE   homology and LIM domains).
GN   NICAL OR MICAL.
OS   Homo sapiens (Human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., SUBCELLULAR LOCATION, TISSUE SPECIFICITY, AND
RP   INTERACTION WITH CASL AND VIMENTIN.
RX   MEDLINE=21964028; PubMed=11827972;
RA   Suzuki T., Nakamoto T., Ogawa S., Seo S., Matsumura T., Tachibana K.,
RA   Morimoto C., Hirai H.;
RT   "MICAL, a novel CasL interacting molecule, associates with vimentin.";
RL   J. Biol. Chem. 277:14933-14941(2002).
CC   -!- FUNCTION: May be a cytoskeletal regulator that connects CasL/HEF1
CC   to intermediate filaments.
CC   -!- SUBUNIT: Associates with the SH3 domain of CasL/HEF1. Interacts
CC   with vimentin.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Expressed in the thymus, lung, spleen, kidney,
CC   testis and hematopoietic cells.
CC   -!- SIMILARITY: Contains 1 calponin-homology (CH) domain.
CC   -!- SIMILARITY: Contains 1 LIM zinc-binding domain.

DR   EMBL; AB048948; BAB86289.1; -.
DR   Genew; HGNC:20619; NICAL.
DR   MIM; 607129; -.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001781; LIM.
DR   Pfam; PF00307; CH; 1.
DR   Pfam; PF00412; LIM; 1.
DR   ProDom; PD000094; LIM; 1.
DR   SMART; SM00033; CH; 1.
DR   SMART; SM00132; LIM; 1.
DR   PROSITE; PS50021; CH; 1.
DR   PROSITE; PS00478; LIM_DOMAIN_1; 1.
DR   PROSITE; PS50023; LIM_DOMAIN_2; 1.
KW   Cytoskeleton; LIM domain; Metal-binding; Zinc.
FT   DOMAIN   508   609   LIM.
FT   DOMAIN   695   757   LIM.
FT   DOMAIN   765   834   PRO-RICH.
SQ   SEQUENCE 1067 AA; 117932 MW; 8A65FD13B72B6DE6 CRC64;

```
Db   347 AlaGlnLeuGluGluMetAlaArgValAlaAlaGlu-----------------ProMet  360
            :::                 |||                                         |||
Qy   961 GCTCGCCACAGACGACGAGATGGCAAGATCCAGGTTACTTTGGAGATCTCTTTGACAGATAC 1020
            |||||||   |||     :::
Db   361 AlaArgAlaArgAspAspAspAlaMetAsnAlaHisLeuLysGluVal----------   376

Qy  1021 GTTCGTATTTCAGATAAAGTAGTGGGCATTCTCATGCGTGCCAGGAAACATGGACTGGTA 1080
            :::     :::|||       |||   |||       :::|||
Db   377 LeuHisAlaAlaAspProMetAlaAsnMetIleGlnIleLysLysArgArgAspThrAlaIle  396

Qy  1081 GACTTTGAAGGAGGAGATGTCTATGGCAAGGCCGAGATGACCATGTTGTG 1128
            |||||||||||:::::::::|||||        |||   |||
Db   397 Asp---ArgGlyGluLeuValTyrProSerTyrHisGlyHisPheVal  411
```

```
Db   901 GlyThrMetAsnAsnTyrProThrTrpArgThrLeuLeuArgArgAlaLysGluGlu   920
Qy   667 CAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCAA---------------   708
             ::::::  |||||                     |||||||||||
Db   921 GluMetLysArgPhe-------------------CysLysAlaGlnThrIleGlnArgArgLeu   935
Qy   708 ---------------   708
Db   936 AsnGluIleGluAlaAlaLeuAlaArgGluLeuGluAlaGluGlyValValLysLeuGluLeuAla   955
Qy   709 -----------CAGAAATATAGCCCAGTGGCAACTTGAAA-------GGGGATGGCAG   750
                        |||                          |||||
Db   956 LeuArgArgGlnSerSerProGlnGlnLysLysTrpValGlyGlnLeuLeu   975
Qy   751 CAGTGGGCTGATGAACAC-----------------------ATACAATCC   777
              |||  |||:::                          |||
Db   976 GlnLeuValAspLysLysAsnSerLeuValAlaGluAlaGluLeuGluMetIleThrVal   995
Qy   778 CAGAAGCTCAATCCTTTCAGTGAAGAGTTGATTACGCGGTCCATGTCCACCCGCCTA   837
              :::::||||||                       ::::|||
Db   996 GlnGluLeuAsn-----------LeuGluGluLysGlnTrpGlnLeuAspGlnGluLeuArg---  1012
Qy   838 CACAAAGGAGATGAGGGCTATGCGCCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCC   897
Db  1013 ----------------GlyTyrMetAsnArgGluGluAsnLeuLysThrAla--AlaAsp  1026
Qy   898 AAGCGTCGCTGAGAGCACATCTACAGGGAAATGATGGAAACATGTGCTTCATTATCTGCACA   957
              :::::|||||||||:     |||||||:::||||||:
Db  1027 ArgGlnAlaAlaGluAspGlnValLeuArgLysLeuValAspLeuVal---------------  1041
Qy   958 ATGGCTCGCCCACAGACGAGACGAAGATGCGAAGATCCAAGGTTACTTTTGGAGATCTCTTTGACAGA  1017
              :::::|||||                      |||
Db  1042 --------AsnGlnArgAspAlaLeuIleArg--------------------PheGlnGlu  1052
Qy  1018 TACGTTCGTATTTCAGATAAAGTAGTGGGC  1047
              |||:::|||:::    ::|||
Db  1053 GluArgArgLeuSerGluLeuAlaLeuGly  1062
```

```
DR   EMBL; D90900; BAA16696.1; -.
DR   PIR; S74544; S74544.
     HAMAP; MF_00100; -; 1.
DR   InterPro; IPR000795; EF_GTPbind.
DR   InterPro; IPR004161; EFTU_D2.
DR   InterPro; IPR000178; IF2.
DR   InterPro; IPR006847; IF2_N.
DR   InterPro; IPR005225; SmaII_GTP.
DR   InterPro; IPR009000; Translat_factor.
DR   Pfam; PF00009; GTP_EFTU; 1.
DR   Pfam; PF03144; GTP_EFTU_D2; 2.
DR   Pfam; PF04760; IF2_N; 2.
DR   PRINTS; PR00315; ELONGATNFCT.
DR   ProDom; PD186100; IF2; 1.
DR   TIGRFAMs; TIGR00487; IF-2; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS01176; IF2; 1.
KW   Initiation factor; Protein biosynthesis; GTP-binding;
KW   Complete proteome.
FT   DOMAIN    496    648    G-DOMAIN.
FT   NP_BIND   502    509    GTP (BY SIMILARITY).
FT   NP_BIND   552    556    GTP (BY SIMILARITY).
FT   NP_BIND   606    609    GTP (BY SIMILARITY).
SQ   SEQUENCE  1001 AA;  108118 MW;  4C7F00DBB56CE31C CRC64;
```

```
Alignment Scores:
Pred. No.:                1.39           Length:        1001
Score:                  116.50           Matches:         81
Percent Similarity:     33.61%           Conservative:    39
Best Local Similarity:  22.69%           Mismatches:     111
Query Match:             5.66%           Indels:         126
DB:                                      Gaps:            18
```

```
US-10-644-659A-1 (1-1146) x IF2_SYNY3 (1-1001)

Qy    12 CGAAAAGGAAAGCGGGAGGCGGCCAGCCAAGAGCGC------------CCTCCGGAAGATACG    62
            |||||
Db    89 LysGlnGluLysSerGlyGlyGlyProSerProAlaArgProThrProProArgProLys    108
Qy    63 CACAGCCACCCTGGTCATCAGCTTGGCCGAGG---TTGGCAGCAGTGGCGAATGAGAA    119
            |||
Db   109 LeuGlnAlaProLysAlaProThrProProGlnProProGlnProValAlaLys-AlaSerAlaPr    128
Qy   120 CAGCATCAGGGAGGCCCCAGGAGCCTACAGGGTGGCTGCCGGGGAGGGACCCAGGAGACCTACACC    179
            :::
Db   128 oLysIleGlnIleLysSerGlnGluGlnGluPro---------AlaGl    138
Qy   180 TCAAGCTCCTAAACCAATCACACCCCCTACT-------------TC    212
            :::||||||||
Db   138 nGluAlaProLysSerValAlaProThrProGlnProLeuAlaProProValProProValProSerE    158
Qy   213 ACACCAGAAGCTTCAGAGTGCCCCAAAGTCGCCA-----------    246
            |||
Db   158 rLeuGlnSerProProSerLysProAlaProProThrProProAlaLysLysAlaAlaPr    178
Qy   247 ----CCCGCCTGCACAGGAACGACATGGAGATGGACAAAGCTCAGAGAGAAAAGCCCCTGAGGT    302
            :::
Db   178 oAlaProArgLeuAlaGlyProProGlyArgThrAlaSerProAsnLysThrAlaValPr    198
Qy   303 TTCTCACATCAAAAGAAAGAAGAGGTGTCCAAAACGTGGTCAGCAAGACTTACGAGAGAGG    362
            :::|:||||
Db   198 oAlaProAlaLysProLysValAsnArgProGluIleValSerLeuLysLysAspAsnArgGl    218
Qy   363 A---------------GGGGACGTGAGCCGCACCTCAGCCACAGTACGAGAGAGGGATGCT--    405
            |||
Db   218 yGlnAlaArgSerProGlyAsp---------ArgGluGluLysValAlaIl    232
Qy   406 -GGTGTGCTTGAACCTGGCACCTGGACAGCGGCAGAATGACAGAGAATGACATTGACAGATCCTTCCACAGCCACGG    464
```

RESULT 9
```
IF2_SYNY3
ID   IF2_SYNY3     STANDARD;     PRT;   1001 AA.
AC   P72689;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Translation initiation factor IF-2.
GN   INFB OR SLR0744.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
```
```
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA   Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA   Yamada M., Yasuda M., Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
```
```
CC   -!- FUNCTION: One of the essential components for the initiation of
CC       protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC       hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC       Also involved in the hydrolysis of GTP during the formation of the
CC       70S ribosomal complex (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: Belongs to the IF-2 family.
```

```
Db 232 eAlaAlaProGluProProLysProLysValGluLeuArgArg-------------Pr 247
         |||||     ::|||||::     :::||
Qy 465 CTCCCAACCGCGAGGAGAAAATGCCAACCTCGTGTCTGAGCTAACCAAGGGCTGGAG 524

Db 247 oLysProProArgProGluGluAspGluAsnLeuProGluLeuLeu---------- 262
          |||        :::   ||||||
Qy 525 AGTGATGGGAGCAGGAGGAGCCCCATGGAGAGTGACAGCGTAGACACAGAGACAGCGG 584

Db 263 ----------GluPheProProLeuSerArgGlyLysGlyValAspGlyAspAsnAspAspAla 279
              ::   ||||||  :::::
Qy 585 CTATGGAGGAGAG------GCTGAGGAGAGCCCCGAGCAGGATGGAGTGCAGGTGGCTGT 638

Db 279 aAspAspAspLeuSerLeuSerThrGluLysProLysPro---------------- 292
          ||:     :::  ||||||:::||
Qy 639 GGTCAGGATCAAGGCGCCCTTGCCCTCCCAGGTAAACAGATTTACAGAGAAACTCAACTG 698

Db 293 ---LysLeuLysArgProThrProProArgLeu---------------------- 302
         :::::||||||||||||||
Qy 699 CAAAGCCCAACAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGATGGCAGCAGTGGGC 758

Db 303 ----------------------GlyLysProAspGlnTrpGl 309
                                |||::
Qy 759 TGATGAACACATACAATCCCAGAAGCTCAAT------------------------- 789

Db 309 uAspAsp-----GluAspGluLysAlaAsnLysAlaLysAlaAlaAsnLysGlyLysAr 327
          |||:        :::  :::||
Qy 790 ----CCTTTCAGTGAAGAGTTGATTACGAGCTGGCAATGTCCACCGCCTACACAAAGG 845

Db 327 gArgProLysMetAspAspAspAspAspLeuAspAspAsp-----------Gl 342
          ||:                     :::
Qy 846 AGATGAGGGC------------------------TATGGCCGGCCC----------- 867

Db 342 yAspAsnGlyProLysProProThrLeuValSerLeuSerIleAlaArgProProLysProLy 362
                                        ||||||:::
Qy 868 ------------AAAGAAGGAACCAAAAACTGCTGAAAGGGCCAAGCGT 903

Db 362 sSerLeuAlaAlaLysProSerThrProThrValAlaLysValLysLys 378
```

RESULT 10
ID   NCR2_MOUSE   STANDARD;   PRT;   2472 AA.
AC   Q9WU42; Q9WU43; Q9WCC1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Nuclear receptor co-repressor 2 (N-CoR2) (Silencing mediator of
DE   retinoic acid and thyroid hormone receptor) (SMRT) (SMRTe) (Thyroid-,
DE   retinoic-acid-receptor-associated co-repressor) (T3 receptor-
DE   associating factor) (TRAC).
GN   NCOR2 OR SMRT.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
RC   TISSUE=Brain, and Spleen;
RX   MEDLINE=99178941; PubMed=10077563;
RA   Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT   "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC   TISSUE=Embryo;
RX   MEDLINE=99199215; PubMed=10097608;
RA   Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
RT   "SMRTe, a silencing mediator for retinoid and thyroid hormone
RT   receptors-extended isoform that is more related to the nuclear
RT   receptor corepressor.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
RN   [3]
RP   INTERACTION WITH HDAC7.

RX   MEDLINE=20107033; PubMed=10640276;
RA   Kao H.-Y., Downes M., Ordentlich P., Evans R.M.;
RT   "Isolation of a novel histone deacetylase reveals that class I and
RT   class II deacetylases promote SMRT-mediated repression.";
RL   Genes Dev. 14:55-66(2000).
CC   -!- FUNCTION: Mediates the transcriptional repression activity of some
CC   nuclear receptors by promoting chromatin condensation, thus
CC   preventing access of the basal transcription.
CC   -!- SUBUNIT: Forms a large corepressor complex that contains SIN3A/B
CC   and histone deacetylases HDAC1 and HDAC2. This complex associates
CC   with the thyroid (TR) and the retinoid acid receptors (RAR) in the
CC   absence of ligand, and may stabilize their interaction with TFIIB.
CC   Interacts with HDAC10 and MINT (By similarity). Interacts with
CC   HDAC7.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=Alpha;
CC   IsoId=Q9WU42-1; Sequence=Displayed;
CC   Name=Beta;
CC   IsoId=Q9WU42-2; Sequence=VSP_003414;
CC   -!- TISSUE SPECIFICITY: Ubiquitous. Also widely expressed in early
CC   embryos.
CC   -!- DOMAIN: The N-terminal region contains repression functions that
CC   are divided into three independent repression domains (RD1, RD2
CC   and RD3). The C-terminal region contains the nuclear receptor
CC   interacting domains that are divided in two separate interaction
CC   domains (ID1 and ID2).
CC   -!- DOMAIN: The two interaction domains (ID) contain a conserved
CC   sequence referred to as the CORNR box. This motif is required and
CC   sufficient to permit binding to unliganded TR and RARS. Sequences
CC   flanking the CORNR box determine nuclear hormone receptor
CC   specificity.
CC   -!- SIMILARITY: Contains 1 SANT-A domain.
CC   -!- SIMILARITY: Contains 1 Myb-like domain.
CC   -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC   FAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF113001; AAD20944.1; -.
DR   EMBL; AF113002; AAD20945.1; -.
DR   EMBL; AF125671; AAD22917.1; -.
DR   TRANSFAC; T04690; -.
DR   MGD; MGI:1337080; Ncor2.
DR   InterPro; IPR001005; Myb_DNA_binding.
DR   Pfam; PF00249; myb_DNA-binding; 2.
DR   SMART; SM00717; SANT; 2.
DR   PROSITE; PS50090; MYB_3; 1.
KW   Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW   Coiled coil; Alternative splicing.
FT   DOMAIN   165   207   COILED COIL (POTENTIAL).
FT   DOMAIN   254   312   INTERACTION WITH SIN3A/B (BY SIMILARITY).
FT   DNA_BIND 429   474   SANT-A (POTENTIAL).
FT   DNA_BIND 609   653   MYB.
FT   DOMAIN   492   560   COILED COIL (POTENTIAL).
FT   DOMAIN   652   682   COILED COIL (POTENTIAL).
FT   DOMAIN   775   804   PRO-RICH.
FT   DOMAIN   989   999   PRO-RICH.
FT   DOMAIN   1351   1357   PRO-RICH.
FT   DOMAIN   2094   2098   CORNR BOX OF ID1.
FT   DOMAIN   2296   2300   CORNR BOX OF ID2.
FT   DOMAIN   494   507   POLY-GLN.
FT   DOMAIN   1615   1619   POLY-ALA.
FT   DOMAIN   2434   2437   POLY-PRO.
FT   VARSPLIC   36   254   Missing (in isoform Beta).
FT   /FTId=VSP_003414.

```
FT   CONFLICT      176    176   M -> RL (IN REF. 2).
FT   CONFLICT      396    402   PPMLYDA -> RHVVRR (IN REF. 2).
FT   CONFLICT      555    555   D -> H (IN REF. 1; AAD20944).
FT   CONFLICT      756    756   T -> M (IN REF. 1; AAD20944).
FT   CONFLICT      785    785   V -> A (IN REF. 2).
FT   CONFLICT      806    846   HHLPHPRLLWTRMNKKPRLLQLPRQRMPRSRSLRPRRSMWE
                                -> PSPRAPPATVDKDEQEARAPAPAPQTEDAKEQKSEAEEI
                                DVG (IN REF. 2).
FT   CONFLICT      856    856   E -> K (IN REF. 1; AAD20945).
FT   CONFLICT      859    859   E -> K (IN REF. 1; AAD20945).
FT   CONFLICT      867    867   E -> K (IN REF. 1; AAD20945).
FT   CONFLICT      895    895   E -> K (IN REF. 1; AAD20945).
FT   CONFLICT      916    916   S -> F (IN REF. 1; AAD20944).
FT   CONFLICT      975    975   I -> IQ (IN REF. 1; AAD20944).
FT   CONFLICT     1046   1063   PKLPTEPPRWSSGLPFPI -> QSYRLSPHAGHRLPSH
                                (IN REF. 2).
FT   CONFLICT     1073   1080   PHAADPSA -> TRADPL (IN REF. 2).
FT   CONFLICT     1133   1133   MISSING (IN REF. 2).
FT   CONFLICT     1149   1149   MISSING (IN REF. 2).
FT   CONFLICT     1157   1157   G -> E (IN REF. 2).
FT   CONFLICT     1172   1201   GSATSGSITKGLPSTRAADGPSYRGSITHG -> APPPVEA
                                SPRASQYPFGCRRPQLQRLYHPR (IN REF. 2).
FT   CONFLICT     1696   1696   A -> S (IN REF. 2).
FT   CONFLICT     1855   1857   MISSING (IN REF. 2).
FT   CONFLICT     1909   1909   A -> P (IN REF. 2).
FT   CONFLICT     1913   1913   A -> G (IN REF. 2).
FT   CONFLICT     1923   1923   G -> A (IN REF. 2).
FT   CONFLICT     1956   1956   N -> S (IN REF. 2).
FT   CONFLICT     1968   1968   A -> G (IN REF. 2).
FT   CONFLICT     2195   2196   TA -> AV (IN REF. 2).
FT   CONFLICT     2213   2214   LE -> SK (IN REF. 2).
FT   CONFLICT     2224   2224   T -> A (IN REF. 2).
SQ   SEQUENCE     2472 AA;  270856 MW;  2A58F4DF7B79285B CRC64;
```

```
Alignment Scores:
Pred. No.:                       2472
Score:            1.78            Length:          2472
Percent Similarity:  115.50      Matches:           82
Best Local Similarity:  37.35%   Conservative:      39
Query Match:      25.31%         Mismatches:       109
DB:               5.61%          Indels:            94
                  1              Gaps:              15
```

US-10-644-659A-1 (1-1146) x NCR2_MOUSE (1-2472)

```
Qy      12  CGAAAAGGAAAGCGGGAGGGGGAGGGCCCAGCCAAGAGCGCCCTTCCGGAAGATACGCACAGCCAC  71
            |||  ||:::: || ::|| :||||||| ::|:
Db     729  ArgValGlyGluCysSerGlyProAlaAlaValAsnAsnSerSerAspThrGluSerVal  748

Qy      72  CCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGCAGAATGAGAACAGATCAGCA  131
                        :||                      :::                  ---Thr  759
Db     749  Pro----------SerProArgSerGluAlaThrLysAsp

Qy     132  GGGCCCAGGAGCCTACAGGCGTGGTCTGCCGCGGAGGACACCCAGGACTCACCTCAAGCTCCTAA  191
            ||||||| :::|||||
Db     760           GlyThrGluAlaLeuProAlaAlaThrGly  774

Qy     192  ACCAATCACACCCCCTACTTCACACCAGAAAGCTCAGAGTGCCCAAAGTCGCCACCCCG  251
            :|||| :|||||
Db     774  nProProValProPro-----GluGluProAlaValAlaProAlaProLeuProSerPr  792

Qy     252  CCTGCCAGAAGGACATGGAGGACAATGGGACAAAGCTCAGAGAAAGCCCCTGAGGTTCTCACAT  311
            :::|||||:: ||                           ::|||||
Db     792  oValProAspAlaSerGly-----------------ProProSerProGluProSerHis--  806

Qy     312  CAAAAGAAAGAGAGGTGTCCAAAACGTGGTCAGCAAGACTTACGAGAGGAGGGACCT  371
            ||||||
Db     806  -----------------------------------------------------------  806

Qy     372  GAGCCACCTCAGCCAC------------------AGGTACGAGAGGGGATGCTGTGTGTGCT  413
            |||   |:                     ::||
Db     807  ---HisLeuProHisProArgLeuLeuTrpThrArgMetAsnLysLysProArgLeuIle  825

Qy     414  TGAACCTGGGCAGCCAGCAGAATGACATTGACAGATCCTCCACAGCCACCGCGCTC------  467
```

```
RA   Grandy D.K., Leduc R., Makam H., Flanagan T., Diliberto E.J.,
RA   Thomas G., Civelli O., Viveros O.H.;
RL   Submitted (OCT-1990) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 634-646.
RC   TISSUE=Adrenal chromaffin;
RX   MEDLINE=9262699; PubMed=7744058;
RA   Strub J.-M., Garcia-Sablone P., Lonning K., Taupenot L., Hubert P.,
RA   van Dorsselaer A., Aunis D., Metz-Boutigue M.-H.;
RT   "Processing of chromogranin B in bovine adrenal medulla.
RT   Identification of secretolytin, the endogenous C-terminal fragment of
RT   residues 614-626 with antibacterial activity.";
RL   Eur. J. Biochem. 229:356-368(1995).
RN   [5]
RP   CHARACTERIZATION OF SECRETOLYTIN.
RX   MEDLINE=96184581; PubMed=8603705;
RA   Strub J.M., Hubert P., Nullans G., Aunis D., Metz-Boutigue M.-H.;
RT   "Antibacterial activity of secretolytin, a chromogranin B-derived
RT   peptide (614-626), is correlated with peptide structure.";
RL   FEBS Lett. 379:273-278(1996).
CC   -!- FUNCTION: Secretogranin I is a neuroendocrine secretory granule
CC       protein, which may be the precursor for other biologically active
CC       peptides. The 16 pairs of basic AA distributed throughout its
CC       sequence may be used as proteolytic cleavage sites.
CC   -!- FUNCTION: Secretolytin has antibacterial activity.
CC   -!- SUBCELLULAR LOCATION: Neuroendocrine and endocrine secretory
CC       granules.
CC   -!- PTM: O-glycosylated (Probable).
CC   -!- SIMILARITY: Belongs to the chromogranin / secretogranin protein
CC       family.
```

```
DR   EMBL; X55027; CAA38846.1; -.
DR   EMBL; U88551; AAC48720.1; -.
DR   EMBL; X55489; CAA39109.1; -.
DR   PIR; S15901; S15901.
DR   InterPro; IPR001819; Chromogranin_AB.
DR   InterPro; IPR001990; Granin.
DR   Pfam; PF01271; Granin; 1.
DR   PRINTS; PR00659; CHROMOGRANIN.
DR   PROSITE; PS00422; GRANINS_1; 1.
DR   PROSITE; PS00423; GRANINS_2; 1.
KW   Sulfation; Cleavage on pair of basic residues; Signal.
FT   SIGNAL        1     20
FT   CHAIN        21    646       SECRETOGRANIN I.
FT   PEPTIDE     418    484       GAWK PEPTIDE.
FT   PEPTIDE     634    646       SECRETOLYTIN.
FT   DISULFID     36     57       BY SIMILARITY.
FT   MOD_RES     158    158       SULFATION (POTENTIAL).
FT   MOD_RES     315    315       SULFATION (BY SIMILARITY).
FT   CONFLICT     64     64       N -> S (IN REF. 1).
FT   CONFLICT     70     70       N -> D (IN REF. 2).
FT   CONFLICT     93     98       SEAPGL -> FRSPRAS (IN REF. 3).
FT   CONFLICT    181    181       T -> M (IN REF. 2).
FT   CONFLICT    261    261       H -> R (IN REF. 2).
FT   CONFLICT    386    386       P -> R (IN REF. 2).
FT   CONFLICT    481    481       H -> L (IN REF. 3).
FT   CONFLICT    597    597       M -> V (IN REF. 2).
SQ   SEQUENCE   646 AA;  73339 MW;  420DB1178F9D9B415 CRC64;
```

```
Alignment Scores:
Pred. No.:                  1.68            Length:          646
Score:                    115.00            Matches:          81
Percent Similarity:       34.99%            Conservative:     53
Best Local Similarity:    21.15%            Mismatches:      121
Query Match:               5.59%            Indels:          128
```

```
DB:                                       1                          Gaps:          21

US-10-644-659A-1 (1-1116) x SG1_BOVIN (1-646)
```

```
Qy    13  GAAAAGGAAAGCGGGGAGGGCCCAGAGC-                                      45
              |||||||              :::  :::
Db   169  GlnGluArgLeuSerGlyGlyProGlyLysAlaGlnThrAlaPheLeuAsnGlnArgAsn  188

Qy    46  ------------GCCCTCCGAAGATACGACAGCCACCCTGGTCATCAGCTTGGCCGCGAGGT  96
                      :::::                              :::
Db   189  GlnThrProAlaLysLysGluGluLeuValSerArgTyrAspThrGlnSerAlaArgGly  208

Qy    97  TGGGCAGCAG---TGGGCGAATGAGAACGACCATCAGCAGGCAGGCCCAGGAGGCCTACA---  147
              ::::::           |||  |||   ::::::::          |||  ||||
Db   209  LeuGluLysSerHisSerArgGlnLysSerGlnGluSerGlyGluGlyGlnThrLysSer  228

Qy   148  -----------GGCTGG                                                  165
                      |||
Db   229  GlnGluAsnTrpProGlnLeuGlnArgHisProGluGlyGlnGluAlaProGlyGlu  248

Qy   166  ACCCAGGAC------TCACCTCAAGCTCCTAACCAATCACACCCCACTTCACACCAG  219
              :::::||          :::::::|||   ::  :::   |||
Db   249  SerGluGluAspAlaSerProGluValAspLysArgHisSerArgPro---ArgHisHis  267

Qy   220  AAAGCTCAGAGTGCCCCAAAG                                             261
              --------------TGGCACCCCGCCGTCGCCAGAA
              :::|||||        |||||||
Db   268  HisGlyArgSerArgProAspArgSerSerGlnGluGlyGlyAsnProProLeuGluGluGlu  287

Qy   262  GGACAT------GGAGATGGACAAAGC--TCAGAGAAAGCCCCTGAGGTTTCTCACATCAAA  315
              |||        ||||||||||||         |||    ||||||||||||    :::::
Db   288  SerHisValGlyThrGlyAsnSerAspGluGluLysAlaArgHisProAlaHisPheArg  307

Qy   316  AAGAAAGAGGTGTCCAAAACGGTGTCAGCAAGACTTACGAGAGACGAGGGGACGTGAGC  375
              -------------------------GluGlyAlaGluTyrGlyGluGluVal--  319
Db   308  AlaLeuGlu-                                                    319

Qy   376  CACCTCAGCCACAGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGCAG---             426
              :::|||||                   ||||||
Db   320  ------------ArgArgHisSerAlaAlaGlnIleAlaProGlyLeuGlnGlyPro  333

Qy   427  --------------------------CCAGAGAAT                                435
              ::: :::
Db   334  AlaArgPheGlyGlyArgGlyArgGluHisGlnAlaLeuArgArgProSerGluGlu  353

Qy   436  GACATTGACAGAATCCTCCACAGCCACGGC---TCCCCAACGCCGGAGGAGAAAATGTGCC  492
              :::::::::        ||||||
Db   354  SerLeuGluGlnGluAsnLysArgHisGlyHisGlyLeuSerProAsp-             367

Qy   493  AACCTGGTGTCTGCGAGCTAACCAAGGGCTGGAGAGTGATGGAGAGCAGGAGGAG---      543
              :::::: :::|||:::             |||||
Db   368  -----------------------LeuAsnMetAlaGlnGlyTyrSerGluGluSerGluGluLysArgGlyPro  384

Qy   544  CCCACATGGAGG---                                                   567
              |||::::: :::|||
Db   385  AlaProGlyProSerTyrArgAlaArgGlyGlyGluAlaAlaAlaTyrSerThrLeuGly  404

Qy   568  GACACAGAGGACAGCGGCTATGGAGGAGGAGGCCCGAGCAGGAGGATGGAGTG  627
              :::::::        ||| ::::::       ||||||
Db   405  GlnThrHisAspGluLysGluPheLeuGlyGluThrHisGluSerGlnArg  424

Qy   628  CAGGTGGCTGGTCGTCAGGATCAAGCGGCCCCTTGCCCTCCCAGGTAAACAGATT--      681
              |||                 ::::::||||||        :::::
Db   425  AspLysAla-                        ArgArgLeuProGlyLeuGluArgAsnTyrLeuAsp  440

Qy   682  --ACAGGAGAAACTCAACTGCAAAGCCCAACAGAAAATATAGCCCAGTGGGCAACTTG  735
              ||||||||              |||:::|||:::::
Db   441  TyrGlyGluGlyLysGlyGluGlyLysAlaAlaArgGlyLysTrpGlnProGlnProGlnAlaAspPro  460

Qy   736  AAAGGG-----------------------AGATGGCAGCAGTGGGCTGGGCTGATGAA  765
              :::                                    :::|||:::||||
Db   461  ArgAspAlaAspGluAsnArgGluGluLysAlaArgLeuLysGlnTyrAlaProHis  480

Qy   766  CACATACAATCCAGAAG------------------CTCAATCCTTTC-----      795
```

```
Db        |||||      :::::         ||||||||||||
      481 HisIleThrGluLysArgLeuGluGluLeuLeuAsnProPheTyrAspProSerGlnTrp 500

Qy    796 ---AGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCCCGCTCACACAAAGGAGATGAG 852
              |||::|          :::            ::|||::|||:::||
Db    501 LysSerSerArgPheGluArgLysAspProMetAspSerPheLeuGluGlyGluGluGlu 520

Qy    853 GGCTATGGC 861
              |||
Db    521 GluAsnGly 523
```

RESULT 12
ID   TRP_DROME   STANDARD;   PRT;   1275 AA.
AC   P19334;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transient receptor potential protein.
GN   TRP.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon-R;
RX   MEDLINE=90180449; PubMed=2516726;
RA   Montell C., Rubin G.M.;
RT   "Molecular characterization of the Drosophila trp locus: a putative
RT   integral membrane protein required for phototransduction.";
RL   Neuron 2:1313-1323(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90148782; PubMed=2482778;
RA   Wong F., Schaefer E.L., Roop B.C., Ally A.H.;
RT   "Overlapping transcription units in the transient receptor potential
RT   locus of Drosophila melanogaster.";
RL   Somat. Cell Mol. Genet. 13:661-669(1987).
RN   [3]
RP   SEQUENCE OF 1126-1275 FROM N.A.
RX   MEDLINE=88042982; PubMed=3118483;
RA   Wong F., Yuh Z.T., Schaefer E.L., Roop B.C., Ally A.H.;
RA   Shao D.;
RT   "Proper function of the Drosophila trp gene product during pupal
RT   development is important for normal visual transduction in the
RT   adult.";
RL   Neuron 3:81-94(1989).
CC   -!- FUNCTION: REQUIRED FOR PHOTOTRANSDUCTION. SUGGESTED TO MEDIATE
CC   CALCIUM ENTRY. SEEMS TO FORM A LIGHT-SENSITIVE CALCIUM PERMEANT
CC   CHANNEL.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC   -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE RHABDOMERIC
CC   MEMBRANES OF THE PHOTORECEPTOR CELLS.
CC   -!- SIMILARITY: Belongs to the transient receptor family. STrpC
CC   subfamily.
CC   -!- SIMILARITY: Contains 2 ANK repeats.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M34394; AAA28976.1; -.
DR   EMBL; M21306; AAA56928.1; -.
DR   EMBL; M18634; AAA28977.1; -.
DR   FlyBase; FBgn0003861; trp.
DR   GO; GO:0016028; C:rhabdomere; IDA.
DR   GO; GO:0015279; F:store-operated calcium channel activity; NAS.
DR   GO; GO:0008377; P:light-induced release of calcium, from inte..; IDA.
DR   GO; GO:0009416; P:response to light; IMP.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR005821; Ion_trans.
DR   InterPro; IPR002153; Trans_receptor.
DR   InterPro; IPR004729; Trp_CaChannel.
DR   Pfam; PF00023; ank; 2.
DR   Pfam; PF00520; ion_trans; 1.
DR   PRINTS; PR01097; TRNSRECEPTRP.
DR   SMART; SM00248; ANK; 2.
DR   TIGRFAMs; TIGR00870; trp; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 1.
DR   PROSITE; PS52297; ANK_REP_REGION; 1.
KW   Ionic channel; Transmembrane; Ion transport; Calcium channel;
KW   Vision; ANK repeat; Repeat.
FT   DOMAIN   1   366   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   367   387   POTENTIAL.
FT   DOMAIN   388   390   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   391   411   POTENTIAL.
FT   DOMAIN   412   418   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   419   439   POTENTIAL.
FT   DOMAIN   440   450   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   451   471   POTENTIAL.
FT   DOMAIN   472   507   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   508   528   POTENTIAL.
FT   DOMAIN   529   541   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   542   562   POTENTIAL.
FT   DOMAIN   563   638   CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   639   659   POTENTIAL.
FT   DOMAIN   660   1275   EXTRACELLULAR (POTENTIAL).
FT   REPEAT   69   98   ANK 1.
FT   REPEAT   143   172   ANK 2.
FT   CONFLICT   285   288   GQRQ -> ASSE (IN REF. 2).
FT   CONFLICT   326   329   RRKQ -> PQE (IN REF. 2).
FT   CONFLICT   365   374   KPFVKFITHS -> NPLSSSSRTP (IN REF. 2).
FT   CONFLICT   785   785   S -> N (IN REF. 2).
SQ   SEQUENCE   1275 AA; 142589 MW; 91CFCDD989689B1 CRC64;

Alignment Scores:
Pred. No.:   Length: 1275
Score: 2.11   Matches: 86
Percent Similarity: 114.00   Conservative: 44
Best Local Similarity: 36.52%   Mismatches: 116
Query Match: 24.16%   Indels: 110
DB: 5.54%   Gaps: 21
                            1

US-10-644-659A-1 (1-1146) x TRP_DROME (1-1275)

Qy   9 GGGCGAAAAGGAAAAGCGGGAGGGCCCAGC------------------------------ 38
              ||::: ||| |||:::||||:::
Db   994 GlyLysProGlyThrMetGlyLysProThrAspAspLysProGlyAspAspLysAsp 1013

Qy   39 CAAGAGGCGCCCTCCGGAAGATACGCACAGCCACCTGGTCATCAGGTTGGCCGGAGGTTG 98
              |||:::|||::::|||:::||:::|::
Db   1014 LysGlnGlnProProLysAspSerLysProSerAlaGly------------GlyProLys--- 1029

Qy   99 GCAGCAGTGGGGCGAATGAGAGAACAGCATCAGGCAGGCCCAGGAGGCCTACAGGCTGGCTGCC 158
                  |||||| :::||
Db   1030 --------------------------------------ProGlyAspGlnLysProThr-Pr 1037

Qy   159 GGGAGGAGGACCCAGGACTCACCTCAAGCTCCTAAACAATCACACCCCTACTTCACACCA 218
              |||||| |||:::|
Db   1037 oGlyAlaGlyAlaProLysProGlnAlaAlaGlyThrIleSerLysProGlyGluSerGl 1057

Qy   219 GAAAGCTCACAGTGCCCCAAAGTCGCCACCCCGCCTGCCAGAAGGACATGGAGAT---- 273
              ||||| ::: |||
Db   1057 nLys---LysAspAlaPro---AlaProProThrLysPro-----------GlyAspThrLy 1072

Qy   274 --------------GGACAAAGCTCAGAGAAAGCCCCTGAGGTTTCTCACATCAA 314
                            |||||| :::
Db   1072 sProAlaAlaProLysProGlyGluSerAlaLys------ProGluAlaAla---AlaLy 1089
```

Qy   315 AAAGAAAGAGGTGTCCAAAACGTGGTCAGCAAGACTTACGAGAGAGGAGGGACGTGAG 374
                |||:::||| ||||||||| |||||||
Db   1089 sLysGluSerLysThrGluAlaSerLysProAlaAlaThrAsnGlyAlaAlaLy 1109

Qy   375 CCACCTCAGCCACAGTACGAGAGGGATGCT--------GGTGTGCTTGAACCTGGGCA 425
                     ::: |||:::||||||
Db   1109 sSerAlaAlaProSerAlaProSerAspAlaLysProAspSerLysLeuLysProGly-- 1128

Qy   426 GCCAGAGAGAATGACAATGACAGAATCCTCCACAGCCACGGCTCCCA-----ACGCGGAG 479
                                   ::: |||:::||| |||:::
Db   1129 -----------------AlaAlaGlyAlaProGluAlaThrLysAl 1138

Qy   480 GAGAAAATGTGCCAACCTGGTCTGCTGAGCTAACCAAGGGCTGGAGAGTGATGGAGCAGGA 539
                |||
Db   1138 aThrAsnGlyAlaSerLysProAspGluLysLysSerGly------------ProGl 1153

Qy   540 GGAGCCC-----------ACATGAGGAGTGACAGCGTGACACAGAGGACACAG 581
                ||||||| ::: |||||| :::||||
Db   1153 uGluProLysLysAlaAlaGlyAspSerLysProGlyAspAlaLysAlaLysAspLysAspLy 1173

Qy   582 CGGCTATGGAGGAGGAGGTGAGGAGGAGGCCCGAGGACGGATGGAGTGCAGGTGGCGTGTGGT 641
                         ::::::||||:::||
Db   1173 sLysProGlyAspApLysApLysApLysProGlyAspAsp------------ 1186

Qy   642 CAGGATCAAGCGGCCCCTTGCCCTCCCAGTAAACAGATTTACAGAGAGAAACTCAACTGCAA 701
                |||:::|||| :::
Db   1187 -LysAspLysLysProAlaAspAsnAsnAspLysLys-------------Pr 1199

Qy   702 AGCCCAACAGAGAAATATAGCCCAGTGGGCAACTTGAAAGGGAGGATGGCAGCAGTGGGCTGA 761
                                  |||::: ||
Db   1199 oAlaAspAspLysAspLysProGlyAspAspLys--------------As 1212

Qy   762 TGAACAACATACAATCCCAGAAGCTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCTGGC 821
                |::| || |||
Db   1212 pLysLysProGlyAspApLysAspLysProSerProSerAspAspLysAsp-LysLys----- 1230

Qy   822 CATGTCCACCCGCCTACACAAAGGAGATGAGGGCTATGGCCCCCAAAGAAGGAACCAA 881
                      ||||:::   ::: ||| ||||||||| |||
Db   1231 ---------ProAlaAspAspLysAspLysProAlaAlaAlaProLeuLys---ProA 1247

Qy   882 AACTGCTGAAAGGGCCAAGCGTGCTG-----------------AGGAGCACATCTA 920
                ::: ||||
Db   1247 laIleLysValGlyLnSerSerAlaAlaAlaGlyGluArgGlyLysSerThrValT 1267

Qy   921 CAGGGAAATGATGATGACATGTGCTTCATTATCTGCACAATGGCTC 964
                ||||||:::   ::: ||
Db   1267 hrGlyArg----------------MetIleSerGlyTrpLeu 1275

RESULT 13
TCOF_HUMAN
ID   TCOF_HUMAN   STANDARD;   PRT;   1411 AA.
AC   Q13428; Q99408; Q99860;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Treacle protein (Treacher Collins syndrome protein).
GN   TCOF1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96154183; PubMed=8563749;
RA   Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA   Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT   "Positional cloning of a gene involved in the pathogenesis of
RT   Treacher Collins syndrome.";
RL   Nat. Genet. 12:130-136(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97228900; PubMed=9074926;
RA   Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,

RA   Dixon M.J.;
RT   "Identification of the complete coding sequence and genomic
RT   organization of the Treacher Collins syndrome gene.";
RL   Genome Res. 7:223-234(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97250498; PubMed=9096354;
RA   Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA   Ashley J.A., Lovett M., Jabs E.W.;
RT   "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
RT   mutations in Treacher Collins syndrome throughout its coding
RT   region.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN   [4]
RP   VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
RP   ARG-53.
RX   MEDLINE=97195537; PubMed=9042910;
RA   Edwards S.J., Gladwin A.J., Dixon M.J.;
RT   "The mutational spectrum in Treacher Collins syndrome reveals a
RT   predominance of mutations that create a premature-termination
RT   codon.";
RL   Am. J. Hum. Genet. 60:515-524(1997).
CC   -!- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
CC   syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
CC   of craniofacial development that occurs with an incidence of
CC   1/50,000 live births. The clinical features of TCS are bilaterally
CC   symmetrical and include: (1) abnormalities of the external ears,
CC   atresia of the external ear canals, and malformation of the middle
CC   ear ossicles, which may result in conductive hearing loss; (2)
CC   lateral downward sloping of palpebral fissures, frequently with
CC   colobomas of the lower eyelids; (3) hypoplasia of the mandible and
CC   zygomatic complex; (4) cleft palate.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U40847; AAC50903.1; -.
DR   EMBL; U76366; AAC51181.1; -.
DR   EMBL; U84664; AAC51185.1; -.
DR   EMBL; U84640; AAC51185.1; JOINED.
DR   EMBL; U84641; AAC51185.1; JOINED.
DR   EMBL; U84642; AAC51185.1; JOINED.
DR   EMBL; U84643; AAC51185.1; JOINED.
DR   EMBL; U84644; AAC51185.1; JOINED.
DR   EMBL; U84645; AAC51185.1; JOINED.
DR   EMBL; U84646; AAC51185.1; JOINED.
DR   EMBL; U84647; AAC51185.1; JOINED.
DR   EMBL; U84648; AAC51185.1; JOINED.
DR   EMBL; U84649; AAC51185.1; JOINED.
DR   EMBL; U84650; AAC51185.1; JOINED.
DR   EMBL; U84651; AAC51185.1; JOINED.
DR   EMBL; U84652; AAC51185.1; JOINED.
DR   EMBL; U84653; AAC51185.1; JOINED.
DR   EMBL; U84654; AAC51185.1; JOINED.
DR   EMBL; U84655; AAC51185.1; JOINED.
DR   EMBL; U84656; AAC51185.1; JOINED.
DR   EMBL; U84657; AAC51185.1; JOINED.
DR   EMBL; U84658; AAC51185.1; JOINED.
DR   EMBL; U84659; AAC51185.1; JOINED.
DR   EMBL; U84660; AAC51185.1; JOINED.
DR   EMBL; U84661; AAC51185.1; JOINED.
DR   EMBL; U84662; AAC51185.1; JOINED.
DR   EMBL; U84663; AAC51185.1; JOINED.
DR   EMBL; U79659; AAB40722.1; -.
DR   EMBL; U79645; AAB40722.1; JOINED.
DR   EMBL; U79646; AAB40722.1; JOINED.
DR   EMBL; U79647; AAB40722.1; JOINED.

```
Db  1150 SerSerProSerValSerSerThrLeuAlaAlaLysAspAspProAspGlyLysGlnGlu 1169
Qy   274 GGACAAGCTCAGAGAAGCCCTGAGGTTTCTCACATCAAAAGAAGAGGTGTCCAAA       333
             :::        |||              :::|||
Db  1170 AlaLysProGlnGlnAlaAlaGlyMetLeuSer-----------ProLysThrGlyGlyLys 1186
Qy   334 ACGGTGGTCAGCAAGACTTACGAGAGGGACGTGAGCCACCTCAGCCACAGTAC          393
                                              |||:::|||
Db  1187 GluAlaAlaSer--------------GlyThrThrProGlnLysSerArgLysPro        1200
Qy   394 GAGAGGAGGATGCTGGT                                              408
             :::       |||||||
Db  1201 LysLysGlyAlaGlyAsnProGlnAlaSerThrLeuAlaLeuGlnSerAsnIleThrGln   1220
Qy   409 --------------------GTGCTGAACCTGGCAGCCAGCAGAATGACATTGACAGA     447
                                   :::   :::
Db  1221 CysLeuLeuGlyGlnProTrpProLeuAsnGluAlaGlnAlaValGlnAlaSerValValLys 1240
Qy   448 ATCCTCCACAGCCACCGGCTCCCAACGCGGAGGAGAAAATGTGCCAACCTGTGTCTGAG     507
             :::::|||
Db  1241 ValLeuThrGluLeuGluLeuGlyGlnGlnGluArgLysValLysValAspThrThrLysGlySer 1260
Qy   508 CTAACCAAGGGCTGG---                                              546
             |||||||
Db  1261 SerArgLysGlyTrpGluSerArgLysLeuSerGlyLysAspGlnProAlaAlaArg     1280
Qy   547 ACATGGAGGAGTGACACGTAGACACAGAGACCAGCGGCTATGGAGGAGAGGCTGAGGAG     606
             |||||||         :::::::||||||
Db  1281 ThrProArgSerLysLysLysLysLysLysLeuGlyAlaGlyLysGluGluGlnAlaSerVal 1300
Qy   607 AGGCCCGAGCAGCAGGATGGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGGCCCCCTGCCCTCC 666
                                          :::
Db  1301 SerProGluLysThrSerThr-------                                    1307
Qy   667 CAGGTAAACAGATTTACAGAGAAACTCAACTGCAAAGCCCACACAGAAATATAGCCCAGTG   726
             |||
Db  1308 --------------ThrSerLysGlyLysAlaLysArgAspLysAlaSerGlyLysAspVal 1322
Qy   727 GGCAACTTGAAAGGAGGATGGCACAGTGGCGTGATGAACACATACAATCCCAGAAGCTC     786
             :::
Db  1323 LysGluLysLysGlyLysLysLysSerLeuGlySer------GlnGlyAlaLysAsp     1338
Qy   787 AATCCTTTCAGTGAAGAGTTGATTACGAGCTCGCCATGTCCACCCGGCTACCACAAAGGA   846
             |||      :::       :::
Db  1339 GluPro--------GluGluLeuGlnLysGlyMetGlyThr---ValGlyGlnGlyGly    1354
Qy   847 GATGAGGGCTATGGCCCGCCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCGTGCT 906
             :::
Db  1355 AspGlnSerAsnProLysSerLysLysLysGlyLysSerAspLysArgLysSerLysLysAsp 1374
Qy   907 GAGGAG                                                          912
             :::|||
Db  1375 LysGlu                                                         1376
```

```
DR   EMBL; U79648; AAB40722.1; JOINED.
DR   EMBL; U79649; AAB40722.1; JOINED.
DR   EMBL; U79650; AAB40722.1; JOINED.
DR   EMBL; U79651; AAB40722.1; JOINED.
DR   EMBL; U79652; AAB40722.1; JOINED.
DR   EMBL; U79653; AAB40722.1; JOINED.
DR   EMBL; U79654; AAB40722.1; JOINED.
DR   EMBL; U79655; AAB40722.1; JOINED.
DR   EMBL; U79656; AAB40722.1; JOINED.
DR   EMBL; U79657; AAB40722.1; JOINED.
DR   EMBL; U79658; AAB40722.1; JOINED.
DR   Genew; HGNC:11654; TCOF1.
DR   MIM; 606847; -.
DR   MIM; 154500; -.
DR   GO; GO:0005730; C:nucleolus; TAS.
DR   GO; GO:0005215; F:transporter activity; TAS.
DR   GO; GO:0001501; P:skeletal development; TAS.
DR   InterPro; IPR006594; LisH.
DR   InterPro; IPR003993; treacle.
DR   Pfam; PF03546; treacle; 3.
DR   PRINTS; PR01503; TREACLE.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Disease mutation; Polymorphism.
FT   DOMAIN        6     38       LISH.
FT   DOMAIN       89     97       POLY-GLU.
FT   DOMAIN      204    207       POLY-SER.
FT   DOMAIN      616    619       POLY-SER.
FT   DOMAIN      919    924       POLY-LYS.
FT   DOMAIN     1285   1289       POLY-LYS.
FT   DOMAIN     1375   1386       POLY-LYS.
FT   DOMAIN     1398   1405       POLY-LYS.
FT   VARIANT      53     53       W -> R (in TCS).
FT                                /FTId=VAR_005630.
FT   VARIANT     439    439       P -> L.
FT                                /FTId=VAR_005631.
FT   VARIANT     810    810       A -> V.
FT                                /FTId=VAR_005632.
FT   VARIANT    1313   1313       A -> V (in dbSNP:15251).
FT                                /FTId=VAR_005633.
FT   VARIANT    1355   1355       D -> G.
FT                                /FTId=VAR_005634.
FT   CONFLICT   1312   1312       K -> Q (IN REF. 2).
SQ   SEQUENCE  1411 AA;  144312 MW;  380203D985C2699 CRC64;
```

```
Alignment Scores:
Pred. No.:                2.13
Score:                  114.00
Percent Similarity:      35.38%
Best Local Similarity:   22.51%
Query Match:              5.54%
DB:                          1
```

```
US-10-644-659A-1 (1-1146) x TCOF_HUMAN (1-1411)

Qy    10 GGCGAAAGAAGGAAAAGCGGGGAGGGCCCA----------GCCAAGAGCGCC----------CTC 51
             ||||||||||||||
Db  1078 GlySerGluGluGlyGlnGluGlyProGlnGlyValAlaLysSerAlaHisThrLeuGlyPro 1097
Qy    52 CGGAAGATACGGCACAGCCACCCTGGTCATCAGCTTGGCCCGAGGTTGGCAGCAGTGGGGG   111
             ||||||||                                  :::::||
Db  1098 ThrProSerArgThrGluThrLeuVal-------------GluGluThrAla           1110
Qy   112 AATGAGAACAAGCATCAGGCAGGCCCAGGAGGCCT----------------ACAGGCTGGCTG 156
             ||:::|||                              :::|||:::::
Db  1111 AlaGluSerSerGluAspAspValAlaProSerGlnSerLeuLeuSerGlyTyrMet      1130
Qy   157 ---CCCGGGAGGGACCCAGGACTCACCTCAAGCTCCTAAACCAATCACACCCCCTACTTCA   213
             :::     ||||||   :::   ||
Db  1131 ThrProGlyLeuThrProAlaAsnSerGlnAlaSerLys---AlaThrProLysLeuAsp   1149
Qy   214 CACCAGAAAGCTCAGAGTGCCCAAAGTCGCCACCCCGCCTGCCCAGAAGGACATGGAGAT   273
             |||:::|||        :::   :::

Length:        1411
Matches:         77
Conservative:    44
Mismatches:     137
Indels:          84
Gaps:            15
```

```
RESULT 14
ID   XPC_DROME      STANDARD;      PRT;   1293 AA.
AC   Q24595;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   DNA-repair protein complementing XP-C cells homolog (Xeroderma
DE   pigmentosum group C complementing protein homolog) (XPCDM) (Mutagen-
DE   sensitive 209 protein).
GN   MUS210 OR XPCC OR XPC.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DP CN BW; TISSUE=Embryo;
```

```
RX   MEDLINE=94173669; PubMed=8127661;
RA   Henning K.A., Peterson C., Legerski R., Friedberg E.C.;
RT   "Cloning the Drosophila homolog of the Xeroderma pigmentosum
RT   complementation group C gene reveals homology between the predicted
RT   human and Drosophila polypeptides and that encoded by the yeast RAD4
RT   gene.";
RL   Nucleic Acids Res. 22:257-261(1994).
CC   -!- FUNCTION: Involved in DNA excision repair. May play a part in DNA
CC       damage recognition and/or in altering chromatin structure to
CC       allow access by damage-processing enzymes (By similarity).
CC   -!- FUNCTION: Involved in nucleotide excision repair of DNA damaged
CC       with UV light, bulky adducts, or cross-linking agents.
CC   -!- SUBUNIT: Heterodimer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: Belongs to the XPC family.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z28622; CAA82262.1; -.
DR   PIR; S42402; S42402.
DR   FlyBase; FBgn0004698; mus210.
DR   InterPro; IPR004583; Rad4.
DR   Pfam; PF03835; Rad4; 1.
DR   TIGRFAMs; TIGR00605; rad4; 1.
KW   DNA repair; DNA-binding; Nuclear protein.
FT   DOMAIN       33     40       POLY-ASP.
FT   DOMAIN      632    637       POLY-SER.
FT   DOMAIN      686    693       POLY-SER.
FT   DOMAIN      922    938       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1195   1211       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     1275   1291       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ   SEQUENCE   1293 AA;  144177 MW;  99DF671F9A4151C5 CRC64;
```

Alignment Scores:
```
Pred. No.:              2.28
Score:                113.50
Percent Similarity:    39.18%
Best Local Similarity: 22.74%
Query Match:            5.52%
DB:                     1
```

US-10-644-659A-1 (1-1146) x XPC_DROME (1-1293)

```
Qy      7 CCGGGCCAAAAAGGAAAAGCGGG------------------------GAGGGCCGAGGAGCGGCC  48
          |||  |||||||||||||||||                          :::  |||||||||||||
Db    551 ProAlaGluLysGluSerSerLysSerThrIleSerLysGluAlaGluLysLysAsnAsn  570

Qy     49 CTCCGGAAGATACGCACAGCACCTGGTCATCAGCTTGGCCCGAGGTTGCGCAGCAGTGG  108
          ::||||                            |||    :::||
Db    571 AlaLysLysAlaGlyAlaLaLysProLeuSerLysSerThrThrLysGly----------  586

Qy    109 GGCAATGAGAACAGCATCAGGCAGCCCAGGAGCCTACAGGCTGCTGCGCGGAGGGACC  168
          :::||:||||||||||||||||||||||||||||  :::|||  |||||
Db    587 ---------------SerGluThrLysSerGlyThrValProLysValLys  599

Qy    169 CAGGACTCACCTCAAGCTCCTAAACCAATCACCATCACCCCTACTTCACCCAGAAAGCTCAG  228
          :::::::|| |||||||||  :::  ||| ::: |||||||||  ||||||||||
Db    600 LysGluLeuSerLeuSerLys---LeuValGluLysSerLys--LeuValGluSerLysHisGlnLysAlaHis  618

Qy    229 AGTGCCCAAAGTCGCCACCCCGCCTGCCAGAAGGACATGGAGAATGGACAAAGCTCA--  285
          :::::: ||||| |||||||
Db    619 ThrSerSerLysSerAspThrSerPheAspGluLysProSerThrSerSerSerLys  638

Qy    286 -------GAGAAAGCCCCTGAGGTTTCTCACATCAAAAAGAGAAGGAGGTGTCCAAAACG  336
          |||:::  ||||||||||        ::: |||
Db    639 CysLeuLysGluGluTyrSerGluLeuGly---LeuSerLysLysLeuLeuLysProThr  657
```

```
Qy    337 GTGGTCAGCAAGACTTACCAGAGAGGAGGGGAGTGAGCCACTCAGCCACAGTGTACGAG  396
          :::  |||||                    |||          |||
Db    658 LeuSerSerLysLeuValLeuLysSerLysAsnGlnSerSerPheSerSerAsn--Lys  676

Qy    397 AGGGATGCTGGTGTGCTTGAACCTGGCCAGAGAATGACATTGACAGAATCCTCCAC  456
          |||||
Db    677 SerAspThrSerPheGluGlu-----AsnProSerThr---------------  687

Qy    457 AGCCACGGCTCCCAACGCGGAGGAGAAAATGCCAACCTGGTGTCTGAGCTAACCAAG  516
          |||      :::  :::  |||  :::|||
Db    688 SerSerSerLysSerLeuLysGluGluThrAlaLysLeuSerSerLysLeuGlu  707

Qy    517 GGCTGGAGAGTGATGGAGCGAGGAGCCCACATGGAGGAGTGACAGC-----------  564
          |||              |||             |||
Db    708 AspLysLysValAlaSerSerAlaGluThrLysThrLysValGlnSerSerLeuLeuLys  727

Qy    565 ---GTAGACACAGAGACAGCGGCTATGGAGGAGGGCTGAGGAGGAGGGCCGAGCAGGAT  621
          |||:::::::::||:::::
Db    728 ArgValThrThrGlnAsnIleSerGluSerGlySerLysGlyAspSerSerLysSer  743

Qy    622 GGAGTGCAGGTGGCTGTGGTCAGGATCAAGCGCCCCTTGCCCTCCCAGGTAAACAGATTT  681
          :::|||||||||||||      ::: |||
Db    744 -----LysValAlaProValAspThrPheSerProValAlaGlyArgThrArgArgAla  761

Qy    682 ACAGAGAAACTCAACTGCAAGCCCAACAAGAAAATAGCCCAGTGGCAAC--  732
          |||                                 :::|||||||:::
Db    762 ThrValLys------ProLysThrGluGluLysProGlnValValGlySerProValIle  779

Qy    733 ---------TTGAAAGGAGATGGCAGCAGTGGGCTGATGAACACATCAATCCCAGAAG  783
          :::  |||  ::::||
Db    780 ProLysLeuMetLeuSerLysValLysGlnLeuAsnAlaLysHisSerAspThrGluAsn  799

Qy    784 CTCAATCCTTCAGTGAAGAGTTTGATTACGAGCTGGCCATGTCCACCGCCTACCACAA  843
          |||                                  ::: |||
Db    800 AlaSerProAlaAsnLysHisLeuGlnGln-----------Arg-AsnThrAr  814

Qy    844 GGAGATGAGGGCTATGGCCGCCCCAAAGAAGGAACCAAAACTGCTGAAAGGGCCAAGCGT  903
          |||              |||:::||:::
Db    814 gGluThrArgSerArgSer----------LysSerProLysValLeuIleSerProSerPh  831

Qy    904 GCTGAGGAGCACATCTACAGGGAAATGATGGACATGTGCTTCATTATCTGCACAATGGCT  963
          |||:::       |||||||
Db    831 eLeuLysLysSerAspGlyAlaAspSerThrSerAlaProGlnLys-HisGlnMetA  851

Qy    964 CGCCAACAGCAGAGATGGCAA---GATCCAGGTTACTTTTGGAGATCTCTT------TGAC  1014
          :::::||||:::|||  ::: |||
Db    851 laProGluThrLysAlaArgIleSerProAsnPheLeuSerGluAlaLeuProAlaArgG  871

Qy   1015 AGATACGTTCG  1025
          |||:::|||
Db    871 lnLeuArgSer  874
```

RESULT 15
```
VSX1_CHICK
ID   VSX1_CHICK              STANDARD;      PRT;    350 AA.
AC   Q9IAL2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Visual system homeobox 1 (Transcription factor VSX1) (Homeobox protein
DE   Chx10-1).
GN   VSX1 OR CHX10-1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20108593; PubMed=10640715;
RA   Chen C.-M.A., Cepko C.L.;
RT   "Expression of Chx10 and Chx10-1 in the developing chicken retina.";
RL   Mech. Dev. 90:293-297(2000).
```

```
CC  -!- FUNCTION: Binds to the 37-bp core of the locus control region
CC      (LCR) of the red/green visual pigment gene cluster. May regulate
CC      the activity of the LCR and the cone opsin genes at earlier stages
CC      of development (By similarity).
CC  -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC  -!- TISSUE SPECIFICITY: Restricted to bipolar cells of the retina and
CC      spinal cord.
CC  -!- DEVELOPMENTAL STAGE: At stage 12, expressed throughout the
CC      invaginating optic vesicles and a subset of cells in the ventral
CC      spinal cord, presumably interneuron precursors. At stage 14, when
CC      the optic cup forms, expression is restricted to a subset of
CC      retinoblasts. At stage 15, expression along the ventral spinal and
CC      hindbrain intensifies and persists beyond stage 20.
CC  -!- SIMILARITY: Belongs to the paired homeobox family.
CC  -!- SIMILARITY: Contains 1 CVC domain.
CC  -!- SIMILARITY: Contains 1 homeobox domain.
CC  ----------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  ----------------------------------------------------------------------
DR  EMBL; AF178670; AAF40312.1; -.
DR  HSSP; P06601; 1FJL.
DR  TRANSFAC; T04141; -.
DR  InterPro; IPR001356; Homeobox.
DR  InterPro; IPR007104; Paired_homeo.
DR  Pfam; PF00046; homeobox; 1.
DR  ProDom; PD000010; Homeobox; 1.
DR  SMART; SM00389; HOX; 1.
DR  PROSITE; PS00027; HOMEOBOX_1; 1.
DR  PROSITE; PS50071; HOMEOBOX_2; 1.
KW  Transcription regulation; Vision; Developmental protein;
KW  Nuclear protein; Homeobox; DNA-binding.
FT  DOMAIN       82     95       GLY/LEU-RICH.
FT  DNA_BIND    157    216       HOMEOBOX.
FT  DOMAIN      217    270       CVC.
FT  DOMAIN       44    127       PRO-RICH.
FT  SITE         52     59       OCTAPEPTIDE MOTIF.
SQ  SEQUENCE   350 AA;  37351 MW;  A60B73B021703C33 CRC64;

Alignment Scores:
Pred. No.:                  2.16       Length:          350
Score:                    113.00       Matches:          69
Percent Similarity:       34.65%       Conservative:     36
Best Local Similarity:    22.77%       Mismatches:       89
Query Match:               5.49%       Indels:          109
DB:                            1       Gaps:             14

US-10-644-659A-1 (1-1146) x VSX1_CHICK (1-350)
```

```
Qy   129  GCAGGCCCAGGAGCCTACAGGCTGGCTGCCGGGAGGACCCAGGACTCACCTCAAGCTCC  188
          ||||||||||||||||||||                        |||||
Db    95  AlaGlyProGlyAlaProCysLeu-LeuPro------------------------AlaPr  106

Qy   189  TAAACCAATCACACCCCTACTTCACCACCAGAAAGCTCAGAGTGCCCAAAGTCGCCACC  248
           || |:::::                          :::  |||
Db   106  oLeuProLeuLeuProAla--------ArgGlyProArgProGlyProPr  123

Qy   249  CCGCCTGCCAGGAGGACATGGAGAT----------------------GGACAAAG  281
           |||::                           ::
Db   123  oProAlaAlaArgArgHisLysGluSerAlaSerAspGluSerLeuSerGlyAspAl  143

Qy   282  CTCAGAGAAAGCCCCTGAGGTTCTCCACATCAAA---AAGAAAGGTGTCCAAAACGGT  338
           ||||||               |||:||| ||||||   ::|||||
Db   143  aSerGluLeuLysMetProAlaSerGlnIleLeuArgLysArgLysArgHisArgVa  163

Qy   339  GGTCAGC----------AAGACTTACGAGAGGAGGGGACGTGAGCCACTTCAGCCA  386
           :::              |:||| |||:  :: ::|||
Db   163  lPheThrAlaHisGlnLeuGluGluLeuGluLysLysAlaPheAsnGluAlaHisTyrProAs  183
```

```
Qy   387  CAGGTACGAGAGGGATGCTGGTGTGCTTGAACCTGGCAGCCAGAGATGACATTGACAG  446
          ||| |||               :::::::
Db   183  pValTyrAlaArgGluMetLeuAlaValLysThrGluLeuProGluAspArgIleGlnVa  203

Qy   447  AATCCTCCAGCACCACGGCTTCCCAACGCGGAGGAGAAAATGTGCCAACCTGGTGTCTGA  506
                                                |||||||
Db   203  lTrpPheGlnAsn---------------------ArgArgAlaLys------  211

Qy   507  GCTAACCAAGGGCTGGAGAGTGATGAGCGAGGAGGAGCCACATGGAGGAGTGCAGCGT  566
                                   |||||     ||||||
Db   212  -------------------TrpArg-------LysArgGluGlyTrpGlyArgSerVa  224

Qy   567  AGACACAGGACAGCGGCTATGAGGAGGAGGCCGAGGAGGGCCCGAGCAGGATGGAGT  626
           ||                                |||
Db   224  lMetAlaGluTyrGlyLeuTyrGly-------------------------------  232

Qy   627  GCAGGTGGCTGCTGTGGTCAGGATCAAGCGCCCTTCCCTCCCTCCAGTA-  672
           |:::::: ||||||:||||||||
Db   233  ------AlaMetValArgHisSerIleProLeuProGluSerIleIleAsnSerAlaLy  250

Qy   673  ------------------------------AACAGATTTACAGAGAA  689
                                          :::::: :::  ::
Db   250  sSerGlyLeuValGlySerCysAlaProTrpLeuLeuGlyMetHisLysLysSerMetGl  270

Qy   690  ACTCAACTGCAAAGCCAACAGAAATATAGCCCAGTGGCAACTTGAAACGGGGAGATGGCA  749
          ::::::: ||||||||:                       ||||
Db   270  uValSerArgLysAlaGluSerGlnGluLysProAlaAspGly-------TrpGl  286

Qy   750  GCAGTGGCTGAT--GAACACATACAA-  779
          |||: ||||||||   ||:|||
Db   286  nAlaGluGlnAspGlyGluGluLeuGlnGlyLysLysGlnAlaSerSerGlnArgSerSerGl  306

Qy   780  GAAGTTCAATCCTTTCAGTGAAGAGTTTGATTACGAGCGGAGCCATGTCCACCCGGCCTACA  839
          :|||:|||||||        ::: :::  :  :|||
Db   306  uIleLysLeuGlyProAlaLysAspProGluAspThrAlaIleAspLeueuSer-------  322

Qy   840  CAAAGGAGGATGAGGGGCTATGGCCGCCCCCAAAGAGGAACCAAAACTGCTGAAAGGGCCAA  899
                                                           :::|||:||:::||
Db   323  -----------------------------------ArgThrAlaLysGlnGluLysGluLy  329

Qy   900  GCGTGCT  906
          |||:::
Db   329  sArgSer  331
```

Search completed: May 4, 2004, 08:52:39
Job time : 38 secs